

FOR OFFICIAL USE ONLY

ACCESS DB #

42000

PLEASE PRINT CLEARLY

Location (Bldg/Room#): CM1-7E15

Mail Box CM1-7E12

Scientific and Technical Information Center

SEARCH REQUEST FORM

m:

Date: 8 May 01 Requester's Full Name: _____ Examiner #: S. DEVI

Art Unit: 1645 Phone (308) 9347 Serial Number: 09/151,409

Results Format Preferred (circle): PAPER DISK E-MAIL

ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): JAMES B. DALE

Earliest Priority Date: 9.12.97

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the cited species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only* Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with appropriate serial number.

Please ask MS. BEVERLY SHEARS to perform this search.

Please see attached claims with key words highlighted and/or Examples and synonyms provided.

Please include the following databases: Embase, Medline, Biosis, CA (Dialog 50), JAPIO, JICTEplus, Dialog 35, 65, 77, 144, 256, 266, 440, 348, 357, 113, 129, 130, 156 and 60.

Please perform an inventor's name search.

Point of Contact:

Beverly Shears

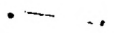
Technical Info. Specialist

CM1 12C14 Tel: 308-4994

Please perform a sequence and an interference search for SEQ ID NO. 15 and 16.

Thank you. ☺

Please return the attached claims and this sequence search request form along with the search reports.



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:42:35 ; Search time 1714.27 Seconds
(without alignments)
9319.598 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatggtgcgcagtag,.....aggcagcgaagtaaaagctt 1158
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13168883 seqs, 6898233199 residues
Total number of hits satisfying chosen parameters: 26337766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1158	100.0	1158	15	US-09-151-409-15	Sequence 15, Appl
2	419	36.2	741	3	US-07-945-954A-2	Sequence 2, Appl
3	328.2	28.3	822	3	US-07-945-954A-3	Sequence 3, Appl
4	301.8	26.1	741	3	US-07-945-954A-6	Sequence 6, Appl
5	301.8	26.1	1029	3	US-07-945-954A-8	Sequence 8, Appl
6	241	20.8	861	3	US-07-945-954A-1	Sequence 1, Appl
7	239.4	20.7	522	3	US-07-945-954A-7	Sequence 7, Appl
8	151	13.0	1332	7	US-08-325-278-5	Sequence 5, Appl
9	124.4	10.7	561	3	US-07-945-954A-4	Sequence 4, Appl
10	82.6	7.1	7218	8	US-07-945-954A-10	Sequence 10, Appl
11	76.2	6.6	408	3	US-07-945-954A-14	Sequence 13, Appl
12	73.2	6.3	1446	18	US-09-404-520-13166	Sequence 2159, Ap
13	72.4	6.3	957	18	US-09-457-659-2159	Sequence 5964, Ap
14	72.4	6.3	937	25	US-09-652-127-5964	Sequence 2729, Ap
15	70.4	6.1	989	17	US-09-397-761A-2729	Sequence 2719, Ap
16	70	6.0	889	17	US-09-397-761A-2719	Sequence 411, App
17	70	6.0	889	25	US-09-641-529-411	Sequence 3082, Ap
18	70	6.0	924	17	US-09-397-761A-3082	Sequence 362, App
19	68	5.9	2943	1	PCT-US99-26796-32	Sequence 2721, Ap
20	67.4	5.8	931	17	US-09-397-761A-2721	Sequence 674, App
21	67.4	5.8	931	25	US-09-641-529-674	Sequence 2728, Ap
22	67.2	5.8	985	17	US-09-397-761A-2728	Sequence 2722, Ap
23	66.4	5.7	857	17	US-09-397-761A-2722	Sequence 444, App
24	66.4	5.7	857	25	US-09-641-529-444	Sequence 13069, A
25	66.2	5.7	948	18	US-09-404-520-13069	Sequence 20050, A
26	66.2	5.7	994	52	US-60-202-214-20050	Sequence 13290, A
27	66.2	5.7	1449	18	US-09-404-520-13290	Sequence 2735, Ap
28	65.4	5.6	979	17	US-09-397-761A-2735	Sequence 2715, Ap
29	65.2	5.6	972	17	US-09-397-761A-2715	Sequence 680, App
30	65.2	5.6	972	25	US-09-641-529-680	Sequence 1953, Ap
31	65.2	5.6	1131	29	US-09-739-449-1953	Sequence 16094, A
32	65	5.6	983	28	US-09-705-926-16094	Sequence 376, App
33	64.4	5.6	786	17	US-09-397-761A-2720	Sequence 232, App
34	64.4	5.6	786	25	US-09-641-529-376	Sequence 13295, A
35	64.2	5.5	3579	1	PCT-US99-26796-232	Sequence 8012, Ap
36	63.8	5.5	1167	18	US-09-404-520-13295	Sequence 1335, Ap
37	63.2	5.5	723	29	US-09-739-449-8012	Sequence 3635, Ap
38	63.2	5.5	892	29	US-09-739-449-1335	Sequence 345, App
39	63.2	5.5	905	25	US-09-652-816-3635	Sequence 354, App
40	63.2	5.5	1998	1	PCT-US99-26796-345	Sequence 13305, A
41	62.8	5.4	834	52	US-60-207-458-354	Sequence 13441, A
42	62.6	5.4	1193	18	US-09-404-520-13305	Sequence 2158, Ap
43	62	5.4	877	51	US-60-196-868-13441	Sequence 5963, Ap
44	61.4	5.3	783	18	US-09-457-659-2158	
45	61.4	5.3	783	25	US-09-652-127-5963	

ALIGNMENTS

RESULT 1
US-09-151-409-15
; Sequence 15, Application US/09151409A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: GROUP A STREPTOCOCCAL VACCINES
; FILE REFERENCE: 481112.410
; CURRENT APPLICATION NUMBER: US/09/151.409A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058.635
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Hexavalent M
; OTHER INFORMATION: fusion gene sequence constructed from
; OTHER INFORMATION: streptococcal type 24, 5, 6, 19, 1 and 3 M protein
; OTHER INFORMATION: DNAs
; NAME/KEY: CDS
; LOCATION: (1)...(1149)
US-09-151-409-15

Query Match 100.0%; Score 1158; DB 15; Length 1158;
Best Local Similarity 100.0%; Pred. No. 2.9e-235;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcatggtcgcgactggtctcagacagatactctggaaagagtaacaaagcgtgct 60
DB 1 gcatgcatggtcgcgactggtctcagacagatactctggaaagagtaacaaagcgtgct 60

QY 61 gacaatttgagatagaaacatacgttaaaacttaagaatagtgacttaagtttaatt 120
DB 61 gacaatttgagatagaaacatacgttaaaacttaagaatagtgacttaagtttaatt 120

QY 121 aataaagcgttaaaagatcataatgatgagtgtaactgaagagtgtagtaattgctaaaag 180
DB 121 aataaagcgttaaaagatcataatgatgagtgtaactgaagagtgtagtaattgctaaaag 180

QY 181 aaactacgtgatccgcgctgactaggggtacataataataaacccagacagacaaagaa 240
DB 181 aaactacgtgatccgcgctgactaggggtacataataataaacccagacagacaaagaa 240

QY 241 gctcttgacaagtatgagctagaaaaccatgacttaaaactaaagaatgaagggttaaaa 300
DB 241 gctcttgacaagtatgagctagaaaaccatgacttaaaactaaagaatgaagggttaaaa 300

QY 301 actgagaatgaagggttaaaactgagaatgaagggttaaaactgagaatgaagggtta 360
DB 301 actgagaatgaagggttaaaactgagaatgaagggttaaaactgagaatgaagggtta 360

QY 361 aaactgagctgacagagtggttcttaggggagcgtgagaaacccgacaaagcacga 420
DB 361 aaactgagctgacagagtggttcttaggggagcgtgagaaacccgacaaagcacga 420

QY 421 gaaacttcaacagtagcgtagagaactctatgtttcaagaactaataatgacaagtta 480
DB 421 gaaacttcaacagtagcgtagagaactctatgtttcaagaactaataatgacaagtta 480

QY 481 ccatgagagtggttatactagccatagcaccagaagataagctaaataattattgac 540
DB 481 ccatgagagtggttatactagccatagcaccagaagataagctaaataattattgac 540

QY 541 gatcttgacgcgaagacatgaattacaacacagaatgagaagttatctctgcagaac 600
DB 541 gatcttgacgcgaagacatgaattacaacacagaatgagaagttatctctgcagaac 600

DB 541 gatcttgacgcgaagacatgaattacaacacagaatgagaagttatctctgcagaac 600
QY 601 ggtgatggttaactcctaggggaagttagaagattctgcagcaaaacaatcccccaataca 660
DB 601 ggtgatggttaactcctaggggaagttagaagattctgcagcaaaacaatcccccaataca 660
QY 661 aatatacgtttacgtcacgaaacaaagggacttaaaagcgagattagagaatgcaatggaa 720
DB 661 aatatacgtttacgtcacgaaacaaagggacttaaaagcgagattagagaatgcaatggaa 720
QY 721 gttgcaggaagagatttttaagagagcgtggtacacctgtttagatcaggttacacaattat 780
DB 721 gttgcaggaagagatttttaagagagcgtggtacacctgtttagatcaggttacacaattat 780
QY 781 actaaacataatagtaattaccacaataataatgcacaagctggcagacttgacctgaga 840
DB 781 actaaacataatagtaattaccacaataataatgcacaagctggcagacttgacctgaga 840
QY 841 caaaagcgtgaatatctaaaaagccttaagtgtggctgagagggctgttacaaaggtta 900
DB 841 caaaagcgtgaatatctaaaaagccttaagtgtggctgagagggctgttacaaaggtta 900
QY 901 aatatacgttcgcgactaggtctcagacagatactcttgaaaaagtcacaagacgtgct 960
DB 901 aatatacgttcgcgactaggtctcagacagatactcttgaaaaagtcacaagacgtgct 960
QY 961 gacaagtttgagatagaaaaacatacgttaaaacttaagaatgtagtacttaagtttaatt 1020
DB 961 gacaagtttgagatagaaaaacatacgttaaaacttaagaatgtagtacttaagtttaatt 1020
QY 1021 aataaagcgttaaaagatcataatgatgagtgtaactgaagagtgtagtaattgctaaaag 1080
DB 1021 aataaagcgttaaaagatcataatgatgagtgtaactgaagagtgtagtaattgctaaaag 1080
QY 1081 aaactacgttaaaatgataatcactctgaaaaagctagtaaaattcaagaattagag 1140
DB 1081 aaactacgttaaaatgataatcactctgaaaaagctagtaaaattcaagaattagag 1140
QY 1141 gcacgttaagtaaaagcgtt 1158
DB 1141 gcacgttaagtaaaagcgtt 1158

RESULT 2
US-07-945-954A-2
; Sequence 2, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945.954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..741
US-07-945-954A-2

Query Match 36.2%; Score 419; DB 3; Length 741;
Best Local Similarity 78.9%; Pred. No. 8.3e-79;
Matches 585; Conservative 0; Mismatches 0; Indels 156; Gaps 1;
QY 7 atgtcgactagctctcagacagactcttgaaaaagtaacagacgctgctgacaag 66
Db 1 ATGCTCGGACTAGGTCTCAGACAGACTCTGGAAAAAGTACAAAGACGCTGCTGACAG 60
QY 67 ttgagatgaacacacgttaacacttaagaatagtagcttaagttaataataaa 126
Db 61 TTGAGATAGAAACAAATACGTTAAACTTAAGATAGTACTGCTTAAATATAA 120
QY 127 gcttaaaagataatgatagtgtaactgaaggttgatgaactgtaaaagaaacta 186
Db 121 GCGTTAAAGATCATAATGATGAGTTAACTGAAGAGTTGAGTAATGCTTAAAGAGAACTA 180
QY 187 cgt----- 189
Db 181 CGTAAATGATAAATCACTATCTGAAAAGCTAGTAAATCAAGATTAGAGGCAGT 240
QY 190 ----- 189
Db 241 AAGGCTGATCTTGAAGAGCAATTAGAGCGCAATGAAATTTTCAACAGCGGATTCAGCT 300
QY 190 -----ggatccgcgctgactaggggt 210
Db 301 AAATCAAAACCTTAGAAGCAGAGAAAGCTGATCTTGAAGGATCGCGCTGACTAGGGCT 360
QY 211 acaataatgacccgcaagagcaaaagctcttgacaagtatgactagagaaacccat 270
Db 361 ACATAATGACCCCAAGCAAGCAAAAGCTCTTGACAAAGTATGAGCTAGAAACCAT 420
QY 271 gacttaaaactaagaatgaagggttaaaactgagaatgaagggttaaaactgagaat 330
Db 421 GACTTAAAACTAAGAATGAAGGTTAAAACTAGAGATGAAGGTTAAAACTGAGNAT 480
QY 331 gaagggttaaaactgagaatgaagggttaaaactgagggtcgacagagtggttccctagg 390
Db 481 GAAGGTTAAAACTGAGAAATGAAGGTTAAAACTGAGGTCGACAGAGTGTTCCTAGG 540
QY 391 gggacggttagaaccgcaacaaagcacgagaaactcttaacagtatgacgtagagaac 450
Db 541 GGGACGGTAGAAACCCGGCAACAGCAGCAACTCTTAAACAGTATGACGTAGAGAAC 600
QY 451 tctatgttaacagctaaataatgacaagttaccatggagagtcggttactactagcacaag 510
Db 601 TCTATGTACAAGCTAAATGACAAGTTACCATGGAGAGTGCCTTACTATAGGCATAGC 660
QY 511 ccagagataagctaaataatattatgaacatttgaacatttgaacgcaaaagacattacaa 570
Db 661 CCAGAGATAGCTAAAAAAATTTATTGACATCTTGACGCAAAAAGAACATGAATTACA 720
QY 571 caacagaatgagaagtatct 591
|||||

Db 721 CAACAGAAATGAGAAATTATCT 741

RESULT 3
US-07-945-954A-3
; Sequence 3, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..822
; US-07-945-954A-3

Query Match 28.3%; Score 328.2; DB 3; Length 822;
Best Local Similarity 87.6%; Pred. No. 1.4e-59;
Matches 404; Conservative 0; Mismatches 3; Indels 54; Gaps 2;
QY 185 tagctggtccgcgctgactaggtgctacataatgacccgcaaaagacagagctc 244
Db 362 TTCAGGATCCCGCTGACTAGGGGTACAAATAATGACCCGCAAGAGCANAAGAGCTC 421
QY 245 ttgacaagtatgactagataaaacccatgacttaaaactaagaatgaagggttaaaactg 304
Db 422 TTGACAAGTATGAGTAGAAAAACCATGACTTAAAACTAAGATGAAGGTTAAAAACTG 481
QY 305 agaatgaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactg 364
Db 482 AGAATGAAGGTTAAAAACTGAGATGAAGGTTAAAACTGAGATGAAGGTTAAAAA 541
QY 365 ctgag-----gtcacagagtggttccctaggggacgg 397
Db 542 CTGAGGTGACCCAGCAACCCAGCTGTTCCTCAGACAGAGTGTTCCTAGGGGACGG 601
QY 398 tagaaacccggacaaagcacgagaaactctttaacaagtatgacgtagagaaactctatg 457

[illegible]

```

RESULT      4
US-07-945-954A-6
; Sequence 6, Application us/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
; US-07-945-954A-6

```

QY 546 tgacgcacaaagaacatgaattacaacacagaaatgaagaagttaattctctgcagcaacggtga 605

Db 63 TGACGCCAAAGAAGAACATGAAATTACAACACAGAAATGAGAAGTTATCTCGATCCAGAGTGT 122

QY 606 tggtaatcctcagggaagttagaagatctctgcagcaacaacatcccgaatcacaaaatat 665

Db 123 TCCTAGGGGCGGTAGAAAACCCGG-----ACAAAGC 155

QY 666 acgtttactgcacgaacaaacaggaacttaaaagcggagattagagaatgcgaatggaagtgc 725

Db 156 ACAGAGAACCTTCTTAACAACGATGACGTAGAGAACCTCTATGTTTCAAGCTTAATAATGACAA 215

QY 726 aggaagagattttaagagagcgttgctactcttgtagatcaggttacacaattataactaa 785

Db 216 CTTAGTCGACCGCGTACTAGGGGTACAATAAATGACCCGCAAGAGCAAAAGAGCTCT 275

QY 786 acataatagtaattaccaacaataatgcacaagctggcgagacttgacctgcagacaaaa 845

Db 276 TGACAAGTATGAGCTAGAAAACCATGACTTAAAAACTAAGAATGAAGGGTTAAAAACTGA 335

QY 846 ggctgaatctaaaaggccttaagtattggctgagaggctgttacagaagattcaatat 905

Db 336 GAATGAAGGGTTAAAAACCTGAGAATCAAGGGTTAAAAACTGAGAAATGAAGGGTTAAAAAC 395

QY 906 cga-----tgtcgcgactcaggtcttcagacagatactctggaaaaagtcacaagaactgc 959

Db 396 TGACCCATGGTTCGCGACTAGGTCTCAGACACATACTCTGGAAAAAGTACAAGAACGTGC 455

QY 960 tgacaagtttgtagatgaacaaacacacgttaaacacttaagaatagtgacttaagtattaa 1019

Db 456 TGACAAGTTTGAGATGAANAACAAATACGTTAAAGCTTAAAGATAGTGACTTAAAGTTTAA 515

QY 1020 taataagcgttaaaagataataatgatgatgagttactgaagctgaagcttgagtaactgaaga 1079

Db 516 TAATAAGCGTTAAAGATCATATGATGATGAGTTACTGACAGTTGAGTAATGCTAAGA 575

QY 1080 gaaactacgtaaaaatgataaatcactatctctgaaaaagctagtaaaattcaagaattaga 1139

Db 576 GAAACTACGTAAAAATGATAAATCACTACTCTGAAAAAGCTAGTAAAAAATCAAGAATTAGA 635

QY 1140 ggcacgtaaataaaagctt 1158

Db 636 GGCACGTAAGCGTGAATCTT 654

RESULT 5

US-07-945-954A-8

Sequence 8, Application US/07945954A

GENERAL INFORMATION:

APPLICANT: Dale, James B.

APPLICANT: Lederer, James W.

TITLE OF INVENTION: Recombinant Multivalent M Protein

TITLE OF INVENTION: Vaccine

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weiser & Associates

STREET: 230 South Fifteenth Street, Suite 500

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/945,954A

FILING DATE: 16-SEP-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Weiser, Gerard J.

REGISTRATION NUMBER: 19,763

Db 244 GCTGATCTT 252

```
RESULT 8
US-08-325-278-5
; Sequence 5, Application US/08325278
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,278
; FILING DATE: 26-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 450023.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1329
; US-08-325-278-5

Query Match 13.0%; Score 151; DB 7; Length 1332;
Best Local Similarity 100.0%; Pred. No. 5.3e-22;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 aacggtgatgttaactctagggaagttatagaagatcttgacgacaaacatcccgcaata 657
|||||
Db 1 AACGGTGATGGTAACTCTAGGGAAGTTATAGAAGATCTTGACGACAAACATCCCGCAATA 60

QY 658 caaatatagcttttcacgcacgaaacaaagacacttaaaagcgagattagagaatgcaatg 717
|||||
Db 61 CAAATATACGTTTACGTCACCAACAAACAGGACTTAAAGCGAGATTAGAGATCAATG 120

QY 718 gaagttcagggaagagattttaagagagctg 748
|||||
Db 121 GAAGTTCCAGGAAGAGATTTTAAGAGAGCTG 151

RESULT 9
US-07-945-954A-4
; Sequence 4, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
```

```
RESULT 7
US-07-945-954A-7
; Sequence 7, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..519
; US-07-945-954A-7

Query Match 20.7%; Score 239.4; DB 3; Length 522;
Best Local Similarity 97.6%; Pred. No. 8.3e-41;
Matches 243; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 910 gtccgactaggtctcagacagatctctggaaaaagtcacaaagcgtgctgcaagttt 969
|||||
Db 4 GTCGCGACTAGGTCTCAGACAGATACCTCTGAAAAAGTACAAAGCCTGCTGACAAGTTT 63

QY 970 gagatagaaaaacatacgtttaaaacttaagaatagtgacttaagttttaataaagcg 1029
|||||
Db 64 GAGATAGAAAAACATACGTTTAAACTTAAGNAATAGTACTTAAGTTTATATAAAGCG 123

QY 1030 ttaaaagatcataatgatgagttaactgaagagttgagtaatgctaaagagaaactacgt 1089
|||||
Db 124 TTAAGAATCATATGATGAGCTTAACCTGAAGAGTTGAGTAATGCTAAAGAGAAACTACGT 183

QY 1090 aaaaatgataactactctcgaagaaagcgtagtaaaattcaagaattagagcagcgtaa 1149
|||||
Db 184 AAAAAATGATAATCACTACTCTCAAAAAAGCTAGTAAAAATCAAGAATTAGAGGCACGTAAG 243

QY 1150 laaaagcgtt 1158
|||
```

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,954A
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372,5672P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-07-945-954A-4

Query Match 10.7%; Score 124.4; DB 3; Length 561;
Best Local Similarity 55.5%; Pred. No. 1.9e-16;
Matches 295; Conservative 0; Mismatches 216; Indels 21; Gaps 2;

Qy 7 atggtcgactaggtctcagacagatctctggaaaaagtaacaagaacgtgctgacaag 66
Dy 1 ATGTCGCGACTAGGTCTCAGACAGATCTCTGAAAAAGTACAGAAAGTCGCGACTAGG 60
Qy 67 ttgagatagaacaacacgttaaaacttaagaatagtgacttaagttttaataataa 126
Dy 61 TCTAGACAGATACTCTGAAAAAGTACAGAAAGTCGCGACTAGTCTCAGACAGATACT 120
Qy 127 gcgttaaaagatcataatgatgagtttaactgaagattgagtaatgcttaagaacta 186
Dy 121 CTGAAAAAGTACAGAAAGTACCGCGTACTAGGGGTACAATAAATGACCCGCAAGA 180
Qy 187 cgtggtacgcgcgtgactaggggttaataatgatcccgcaagaagcaaaagagctctt 246
Dy 181 GCAAAAAGACGCGTACTAGGGGTACAATAAATGACCCGCAAGAAGCAAAAGACCGGTG 240
Qy 247 gacaagtatgagtgaataaaccatgaacttaaaactaagaatgaagggttaaaactgag 306
Dy 241 ACTAGGGGTACATAAATGACCCGCAAGCAAAAGTACAGAAAGTTCCTTAGG 300
Qy 307 aatgaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaact 366
Dy 301 GGGACGGTAGAAACCCGCAAGACAGAGAGTGTTCCTAGGGGACGGTAGAGAAAC 360
Qy 367 gaggttcgac-----agagttttccctagggggcgtagaataacccggacaaagcaga 420
Dy 361 CGGCACAAAGCAGGAAGAGTGTTCCTAGGGGACGGTAGAGAAACCCGGACCAAGCAGA 420

Qy 421 gaacttttaacaagtatgacgtagagaaactctatgttacaagctaatagcaagtta 480
Dy 421 CCA-----TGGAGAGTGGCTTATATAGTCATACGCCAAGATAAGCTA 465
Qy 481 ccatggagagtgctgtatatactaggcatagccagagaagataagctaaacaaaaa 532
Dy 466 AAAAAAAGAGTGGTATATAGTCATACGCCAAGATAAGCTAAAAAAA 517

RESULT 10
US-08-466-194-14/C
Sequence 14, Application US/08466194
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, NW, Suite 500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,194
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,463
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/935,313
FILING DATE: 26-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/201 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-466-194-14

Query Match 7.1%; Score 82.6; DB 8; Length 7218;
Best Local Similarity 8.9%; Pred. No. 2.3e-07;
Matches 40; Conservative 239; Mismatches 168; Indels 0; Gaps 0;

Qy 115 tttaataaagcgtttaaagaatcataatgagtaactgaagtgatgagtaagttgagtaatgct 174
Dy 1511 TTTCAAAAACGGCATGTAGGCATCACTGTAATACCTATCTATGCAAGTAGTTAAAGAG 1452
Qy 175 aaagagaactacgttgatccgctgactaggggtacaataaatgaccgcaaaagaga 234
Dy 1451 ATAGAAGAATTGCTACRR 1392

QY 235 aaagaagctcttgacagatgagctagagaaacacatgacttaaaaaaactaagaatgaagg 294
Db 1391 RRR 1332
QY 295 ttaaaaactgagaatgaagggttaaaaactgagaatgaagggttaaaaactgagaatgaa 354
Db 1331 RRR 1272
QY 355 ggggttaaaaactgagctgacagagtgcttctaggggacggtagaaaaacccggacaaa 414
Db 1271 RRR 1212
QY 415 gcaagagaactcttaacaagatgacgtagagaactctatgtttacaaagctaatgac 474
Db 1211 RRR 1152
QY 475 aagttaccatgagagctgcttactactaggtacgcacagacagataagctaaaaaaatt 534
Db 1151 RRR 1092
QY 535 atgacgatctgacgcaaaagaacat 561
Db 1091 RRR 1065

RESULT 11
US-07-945-954A-10
; Sequence 10, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; US-07-945-954A-10

Query Match 6.6%; Score 76.2; DB 3; Length 408;
Best Local Similarity 71.5%; Pred. No. 2.9e-06;
Matches 118; Conservative 0; Mismatches 38; Indels 9; Gaps 1;
QY 480 accatggagagtgcttactactagctacgacagataaactaaaaaattattga 539
Db 150 ACCATGGAGAGTGGTTTACTAGGCATACCCAGAAGATAAGCTAAAAAACTGCAGGA 209
QY 540 cyatgttgacgcaaaagaacatgaattacaacaacagaalgaagaagttatctctgcagaa 599
Db 210 TGCTAGGAGTCTTAATGAGAGAGTTTCTTAGACATGTTAA -----ATAATCGATAA 260
QY 600 cagtgatgtaatcctaggaagttatagaagatctctgcagaaa 644
Db 261 CGGTGATGTTATCTTAGGAGAGTTATAGAAGATCTTGCAGCAGA 305

RESULT 12
US-09-404-520-13166
; Sequence 13166, Application US/09404520
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Ghodssi, Azita
; APPLICANT: Hinkie, Gregory J.
; APPLICANT: McIninch, James
; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 13166
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
; FEATURE:
; OTHER INFORMATION: unsure at all n locations
; US-09-404-520-13166
Query Match 6.3%; Score 73.2; DB 18; Length 1446;
Best Local Similarity 44.0%; Pred. No. 1.6e-05;
Matches 493; Conservative 0; Mismatches 618; Indels 9; Gaps 4;
QY 41 aaaagtacaaagacgctgacaaagtttgagataaaaacaacacgtttaaacttaaga 100
Db 285 aaaaataaaaaataaataaagataaagaaaaataaagataaataaataaataaataa 344
QY 101 a--tagtgacttaagttttaataataaagcgttaaaagatcataatgatgagttactga 158
Db 345 aattatagaaaaaataaataaataaataaataaataaataaataaataaataaataa 404
QY 159 agagttgagtaatgctaaagagaactacgtggtgacgcgctgactaggggtacataaa 218
Db 405 aaaaaataaagagggaagaaaaataaagagaaaaataaagaaaaataaagaaaaataa 464
QY 219 tgaccgcgaagagcaaaagagctcttgacaagtatgagctagagaaacccatgactaaa 278
Db 465 taaaagaaaaaataaataaataaataaataaataaataaataaataaataaataaataa 524
QY 279 aactaagaatgaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggtt 338
Db 525 atataagaaaaaagggttaaaaaataaataaataaataaataaataaataaagaaaaata 584
QY 339 aaaaactgagaatgaagggtttaaactgaggtgcgacagaggtttctctctggggacggt 398
Db 585 aaattataaaataaataatagataaaataaagaaaaagaaatgaatgataaaaaataa 644
QY 399 agaaaaacccggacaaagcagcgaactcttaacaaagtatgacgtagagaactatgttt 458
Db 645 ataaagaagagaaaaaagaaaaataaataaataaataaataaataaataaataaataa 704

QY	459	acaagctaaatgacaagttaccatggagagtgcgttatactactaggaacacccagaaga	518
Db	705	aagattaaaaaaacaaaaaaagataagataataataataataataataataat	764
QY	519	taagctaaaaaaattattgagatcttgagcgcacaaagaacatgaattacaacacgaa	578
Db	765	aaatgagaatttatagaaaaaaataagagataaaaaaaagataaataatggagaaa	824
QY	579	tgagaagtattctctgcagaacgggatg-gtaactcttaggaagttatagaagatcttg	637
Db	825	taagaaataatagataataaaaaatgaagaaaaataaataagatatgatataaaaaata	884
QY	638	cagcaaacacatccgcacatacaaaaatacgttttcagtcacgaaaaacaggactaaaag	697
Db	885	gtaatatagataataatacagagagaaaaaagaataatgagaaaaatagagaaaaaa	944
QY	698	cgagattgagaatgcgaatgggaagttgcaggaagagatttaagagagctggctctgt	757
Db	945	aaaaattaaagaaaataaaaaaaataaaaaaaataagtaaaaaaagaaaagatat	1004
QY	758	tagatcaggtttacacattatatactaaacabaa----tagtaattaccacacataaat	813
Db	1005	agaataaataataatgaaaaaagaataaaaataagagaagaataataatgataaaaa	1064
QY	814	gcacagctggcagacttgacctgagcagacaaaggctgaatatctaaaggccctaatgat	873
Db	1065	aaaaaaaaaaaaaagtgaataaaaaataataataataataataataataataa	1124
QY	874	tgggctgagaggctgttacagaggttaatatcgatgctgcagctaggtctcagacagat	933
Db	1125	aagaaataataatgaaaaaaaaggtaaaaaataaagaaaaaagaataataaaaaaat	1184
QY	934	actctgaaaaaagtcacagaacgctgcagacagtttga--gatagaaaaaatacagttaa	991
Db	1185	agtaaaaaataataaaaaaaatacaaaaaaattataaagaagaagtaaatagtaaa	1244
QY	992	aacttaagaatgctgaacttaagttttaataataaagcgttaaaagatcataatgtagct	1051
Db	1245	ataaaaaataatataagagaataataatacacagtgagaataaataagataaaaaata	1304
QY	1052	taactgaagagttgagtaagtctaaagagaacactacgtataaaatgatataactactctg	1111
Db	1305	aataataagattaataaaaaaataataatagatatagattgagataagaaaaaa	1364
QY	1112	aaaaagctagtaaaattcaagaattagagcagtaagta	1151
Db	1365	atagaagaagggaagtaataatgaaattataaaacta	1404

```

RESULT 13
US-09-457-659-2159
; Sequence 2159, Application US/09457659
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Comrack, Christopher
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM
; TITLE OF INVENTION: HUMAN NATURAL KILLER CELL AND HUMAN BRONCHIAL EPITHELIUM
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: 1600.1071001
; CURRENT APPLICATION NUMBER: US/09/457,659
; CURRENT FILING DATE: 1999-12-09
; EARLIER APPLICATION NUMBER: 60/111,677
; EARLIER FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/113,891
; EARLIER FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 2259
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2159
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Homo sapiens

```

[illegible]

Search completed: May 13, 2001, 05:29:32
Job time: 2817 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:42:05 ; Search time 88.15 Seconds
(without alignments)
4928.574 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacatggtgcgactgagctg.....aggcagcgaagtaaaagctt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262727 seqs, 187588001 residues

Total number of hits satisfying chosen parameters: 525454

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA_New.*

- 1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.2	5.6	1131	5	US-09-739-449-1953
2	63.2	5.5	723	5	US-09-739-449-8012
3	63.2	5.5	892	5	US-09-739-449-1335
4	59.6	5.1	1250	5	US-09-739-449-1859
5	58.2	5.0	43546	6	US-60-254-168-38
6	56.6	4.9	798	5	US-09-739-449-2641
7	56.4	4.9	909	5	US-09-739-449-636
8	54.6	4.7	543	5	US-09-739-449-8029
9	53.4	4.6	834	5	US-09-739-449-2256
10	52.4	4.5	902	5	US-09-739-449-1929
11	52	4.5	1167	5	US-09-739-449-2034
12	50.8	4.4	1040	5	US-09-739-449-2287
13	50.8	4.4	147068	6	US-60-248-505-357
14	50.8	4.4	334854	6	US-60-248-505-28
15	50.6	4.4	937	5	US-09-739-449-2534
16	50.4	4.4	845	5	US-09-801-833-4847
17	50	4.4	1006	5	US-09-739-449-621
18	48.8	4.2	780	5	US-09-739-449-778
19	48.4	4.2	757	5	US-09-739-449-2286
20	47.8	4.1	753	5	US-09-739-449-1234
21	47	4.1	944	5	US-09-739-449-492
22	46.8	4.0	107487	6	US-60-248-505-231
23	46	4.0	787	5	US-09-739-449-2501
24	46	4.0	1418	5	US-09-739-449-1613
25	44.8	3.9	684	5	US-09-801-833-5354
26	44.2	3.8	499	5	US-09-739-449-1332
27	44	3.8	2795	1	PCT-US01-04098A-2166

28	44	3.8	3058	1	PCT-US01-04098A-198	Sequence 198, App
c 29	43.8	3.8	776	5	US-09-739-449-2257	Sequence 2257, App
c 30	43.8	3.8	4017	5	US-09-783-514-1895	Sequence 1895, App
31	43.6	3.8	4704	5	US-09-801-833-7699	Sequence 7699, App
c 32	43	3.7	681	5	US-09-739-449-2434	Sequence 2434, App
c 33	42.8	3.7	778	5	US-09-739-449-2435	Sequence 2435, App
c 34	42.6	3.7	345	5	US-09-739-449-8021	Sequence 8021, App
c 35	42.4	3.7	3491	4	US-08-956-171C-117	Sequence 117, App
36	42.2	3.6	987	5	US-09-739-449-719	Sequence 719, App
c 37	42	3.6	629	5	US-09-739-449-1185	Sequence 1185, App
c 38	42	3.6	988	5	US-09-739-449-2009	Sequence 2009, App
39	42	3.6	16596	1	PCT-US01-01351-504	Sequence 504, App
40	41.8	3.6	719	5	US-09-739-449-2589	Sequence 2589, App
41	41.8	3.6	1181	5	US-09-739-449-273	Sequence 273, App
c 42	41.8	3.6	5768	1	PCT-US01-01339-8019	Sequence 8019, App
c 43	41.8	3.6	5768	1	PCT-US01-01329-2852	Sequence 2852, App
44	41.6	3.6	315	5	US-09-739-449-366	Sequence 366, App
45	41.6	3.6	548	5	US-09-739-449-1457	Sequence 1457, App

ALIGNMENTS

RESULT 1
US-09-739-449-1953
; Sequence 1953, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 1953
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1131)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-1953

Query Match	5.6%	Score 65.2;	DB 5;	Length 1131;
Best Local Similarity	41.6%	Pred. No. 2.4e-05;		
Matches 342;	Conservative	0;	Mismatches 479;	Indels 1;
Gaps	1;			
QY	41	aaaaagtacaaagcgtgctgacaaagtttgagatagataaaacatacgttataaacttaaga	100	
Db	144	aaaaaaaaaaganaaaagaaaaaagaaaaaagaaaaaannnnnnnaaaaaa	203	
QY	101	atagtacttaagttttaataataaagcgttaaaagctataagtcataatgatgtaactgaag	160	
Db	204	aaaaanaaaaaaaggggaaaaaagaaaaaagaaaaaagaaaaaannnnnaaaangaa	263	
QY	161	agtgagtaatgttaagagaaactactggtggtcccgtagctaggggtacataataatg	220	
Db	264	aaaaaaaaaagaaaaaagaaaaaagaaaaaaggggagaaaaaagaaaaa	323	
QY	221	accgcgaagagcaaaagagctcttgacaaagtatagctagataaaacccatgacttaaaa	280	
Db	324	annnaaaaaaagaaaaaagaaaaaagaaaaaannnnnaaaaaaagaaaaa	383	
QY	281	ctaagatgaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaa	340	
Db	384	aaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaaannnnnaaaaaa	443	
QY	341	aaactgagaatgaagggttaaaactgagtcagacagagtgttctcctaggggacggtag	400	

```
Db 444 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaa 503
Qy 401 aaaaaccgagacaaagcaggaacttcttaacagtagtgcgtagagaaactctgttac 460
Db 504 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 563
Qy 461 aagctaaatgacaaagttaccatgagagtcggttactactaggcatcgcagaagata 520
Db 564 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 623
Qy 521 agctaaaaaattattgacgactcttcgcgcaaaagagacatgaattacaacaacagatg 580
Db 624 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 683
Qy 581 agaagttactctgcagaaagcgtggtgtaactctcagggaaggttatagaagattctgcag 640
Db 684 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 743
Qy 641 caaacaatccgcgaatacaaaaataatcgtttacgtcacgaaaaacaag-gacttaaaagcg 699
Db 744 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 803
Qy 700 agattagagaatgcgaatggaagttgcaggaagagagattttaagagagcgtggtacctgttta 759
Db 804 aaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 863
Qy 760 gatcaggttacacaattatatactaacaataatagtaattaccacaataataatgcacaa 819
Db 864 gnaaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 923
Qy 820 gctggcagacttgacctgagacaaaaggctgaatatctaaaaa 861
Db 924 aaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaaaagagaaaa 965

RESULT 2
US-09-739-449-8012
; Sequence 8012, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 8012
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(723)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-8012

Query Match 5.5%; Score 63.2; DB 5; Length 723;
Best Local Similarity 35.5%; Pred. No. 5.9e-05;
Matches 182; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

Qy 642 aaaaatccgcgaatacaaaaataatcgtttacgtcacgaaaaaagagacttaaaagcgag 701
Db 31 aaaaaannnnnaaaaaaagagaaaaaannnnnaaaaaaaggggagngagaagn 90
Qy 702 attagagaatgcgaatggaagttcaggaagagattttaagagagcgtggtacctgttaga 761
Db 91 aaggaagaaaaaagagaaagtcgaggaagagatttttaagagagcgtggtacctgttaga 761
Qy 762 tcagggttacacaattatatactaacaataatagtagtaattaccacaataataatgcacagc 821
Db 94 aaggaagaaaaaagagaaagannnnnaaaaaaaggggggagngagannnaana 150
Qy 762 tcagggttacacaattatatactaacaataatagtagtaattaccacaataataatgcacagc 821
```

```
Db 151 nnaaannnnnaaaaaaagagaaaaaannnaannnaaaaaaannnaaaaa 210
Qy 822 tgcgagacttgacctgagacaaagcgtgaatatctaaagccttaagtattgggctga 881
Db 211 aannnnnaaaaaaannnnnaannnnnaaaaaaagagaaaaaannnnnaaaaa 270
Qy 882 gagcgctgttaacagagttaaatatcgtatgcgcgactaggtctcagacagatcactctgga 941
Db 271 aaaaaaagagaaaaaannnaaaaaaagagaaaaaannnnnaaaaaaann 330
Qy 942 aaaaagtaacagagacgctgctgacagttgttgagtagagaaacaatactgttaaaacttaagaa 1001
Db 331 anaaaaaagagaaaaaannnaaaaaaagagaaaaaannnnnaaaaaaann 390
Qy 1002 tagtgacttaagttttaataataaagcgtttaaagatcataatgatgatgtaactgaaga 1061
Db 391 aannnnnaaaaaaannnaaaaaaagagaaaaaannnnnaaaaaaannnaaa 450
Qy 1062 gttgagtaatgtaaaagagaaacttacgtaaaaatgataatcactctctgtaaaaaagctag 1121
Db 451 aaaaaaagagaaaaaannnnnaaaaaaagagaaaaaannnnnaaaaaaann 510
Qy 1122 taaaattcagaagattagagcagcagtaagtaaaa 1154
Db 511 aaaaaaagagaaaaaannnnnaaaaaaannnnnaaa 543

RESULT 3
US-09-739-449-1335
; Sequence 1335, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 1335
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(892)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-1335

Query Match 5.5%; Score 63.2; DB 5; Length 892;
Best Local Similarity 35.5%; Pred. No. 6.1e-05;
Matches 182; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

Qy 642 aaaaatccgcgaatacaaaaataatcgtttacgtcacgaaaaaagagacttaaaagcgag 701
Db 199 aaaaaannnnnaaaaaaagagaaaaaannnnnaaaaaaaggggagngagaagn 258
Qy 702 attagagaatgcgaatggaagttcaggaagagattttaagagagcgtggtacctgttaga 761
Db 259 aaggaagaaaaaagagaaagannnnnaaaaaaaggggagngagannnaana 318
Qy 762 tcagggttacacaattatatactaacaataatagtagtaattaccacaataataatgcacagc 821
Db 319 nnaaannnnnaaaaaaagagaaaaaannnnnaaaaaaaggggagngagaagn 378
Qy 822 tgcgagacttgacctgagacaaagcgtgaatatctaaaggccttaagtattggcctga 881
Db 379 aannnnnaaaaaaannnnnaaaaaaagagaaaaaannnnnaaaaaaannnaana 438
Qy 882 gagcgctgttaacagagtttaataatcgtatgcgcgactaggtctcagacagatcactctgga 941
```


Search completed: May 13, 2001, 05:03:55
Job time: 1310 sec

	Query Match	4.4%	Score 50.6;	DB 5;	Length 937;
	Best Local Similarity	44.1%;	Pred. No. 0.027;		
	Matches 212;	Conservative	0;	Mismatches 269;	Indels 0;
				Gaps	
Qy	654	aatacaaaataacatgttttacgtcacgaacaacagagctt	aaaaagcagagattagagaatgc	713	
Db	935	AAAGAAAAAGAAAAAAGAAAAAATTTAAAGAAAAAGAAAT	CAAAAAAAGTAA	876	
Qy	714	aatggaagtgcggaagagattttaaagagagcttggtac	cttgccttgtagtcaggttacaca	773	
Db	875	AAATAAAAAANAANAATTAATAATAGAAAAANAANA	AGAAAGAAATTA	816	
Qy	774	attatatactaaacataatagtaattaccaacaataata	tcacaaagctggcgagacttga	833	
Db	815	AATAAAAAANAANAATTAATAATAATAATAATAATA	AAAAAAGAAAGATGAAAAAATAA	756	
Qy	834	ccitgaacaaaagcgtgaatctctaaagagccctta	atgatctggcctgagagcgtgttaca	893	
Db	755	AAAAAGAAAAANAATTAANAAGAAAAAGAAAAAG	AGTATAAAAAAAGAAAGAAAAA	696	
Qy	894	agngttaaatactcgatgctgcgactaggtcttcagacag	atactctggaaagaagtcacaaga	953	
Db	695	AAAAATTAAGAAAAAANAANAATGTTTAAAAAANA	ATTAAGATAAATAAGAAAAAANAANAATAAAAAAT	636	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 05:48:01 ; Search time 1719.94 Seconds
(without alignments)
9288.875 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcatcgatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13168883 seqs, 6898233199 residues

Word size : 10

Total number of hits satisfying chosen parameters: 6066217

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending_Patents_NA_Main.*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US098_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US099_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US099G_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US099H_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US099I_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US099J_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US099K_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US099L_COMB.seq.*

Result No.	Score	Query Match	Length	ID	Description
1	1158	100.0	1158	15	US-09-151-409-15
2	402	34.7	741	3	US-07-945-954A-2
3	240	20.7	861	3	US-07-945-954A-1
4	217	18.7	522	3	US-07-945-954A-7
5	217	18.7	741	3	US-07-945-954A-6
6	217	18.7	822	3	US-07-945-954A-3
7	217	18.7	1029	3	US-07-945-954A-8
8	151	13.0	1332	7	US-08-325-278-5
9	53	4.6	561	3	US-07-945-954A-4
10	52	4.5	204	3	US-07-945-954A-9
11	52	4.5	408	3	US-07-945-954A-10
12	52	4.5	918	3	US-07-945-954A-5
13	48	4.1	765	3	US-07-945-860A-2
14	48	4.1	765	3	US-07-945-860B-2
15	48	4.1	765	13	US-08-914-479-3
16	48	4.1	855	3	US-07-945-860A-3
17	48	4.1	855	3	US-07-945-860B-3
18	48	4.1	855	13	US-08-914-479-5
19	36	3.1	417	3	US-07-945-860A-1
20	36	3.1	417	3	US-07-945-860B-1
21	36	3.1	417	13	US-08-914-479-1
22	27	2.3	33	15	US-09-151-409-14
23	25	2.2	30	15	US-09-151-409-5
24	25	2.2	36	15	US-09-151-409-1
25	24	2.1	30	15	US-09-151-409-2
26	24	2.1	30	15	US-09-151-409-3
27	24	2.1	30	15	US-09-151-409-4
28	24	2.1	30	15	US-09-151-409-6
29	24	2.1	30	15	US-09-151-409-7
30	24	2.1	30	15	US-09-151-409-8
31	24	2.1	30	15	US-09-151-409-9
32	24	2.1	30	15	US-09-151-409-10
33	24	2.1	30	15	US-09-151-409-11
34	24	2.1	30	15	US-09-151-409-12
35	24	2.1	30	15	US-09-151-409-13
36	23	2.0	37066	56	US-60-243-468-190
37	22	1.9	32	14	US-09-095-028-3
38	21	1.8	351	22	US-09-572-409-39350
39	21	1.8	30570	28	US-09-702-134-29869
40	21	1.8	30763	24	US-09-620-392-11277
41	20	1.7	286	24	US-09-637-086A-35904
42	20	1.7	405	19	US-09-521-640-265461
43	20	1.7	405	46	US-60-140-769-50892
44	20	1.7	612	25	US-09-654-617-9386
45	20	1.7	612	27	US-09-684-016-9386
46	20	1.7	1524	25	US-09-654-617-194272
47	20	1.7	1524	27	US-09-684-016-194272

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

us-09-151-409-15.oligo.rnmpm

Tue May 15 07:26:54 2001

48	20	1.7	8611	18	US-09-404-520-5819	Sequence 5819, Ap	c 121	19	1.6	4417	27	US-09-698-012-8398	Sequence 8398, Ap
49	20	1.7	32768	52	US-60-200-382-31	Sequence 31, App1	c 122	19	1.6	4417	27	US-09-698-014-5331	Sequence 5331, Ap
50	20	1.7	32768	52	US-60-200-382-32	Sequence 32, App1	c 123	19	1.6	4417	27	US-09-699-997-11577	Sequence 11577, A
51	20	1.7	32768	52	US-60-200-382-33	Sequence 33, App1	c 124	19	1.6	4417	27	US-09-699-998-9374	Sequence 9374, Ap
52	20	1.7	32768	53	US-60-212-659-191	Sequence 191, App	c 125	19	1.6	4417	28	US-09-710-285-2019	Sequence 2019, Ap
53	20	1.7	32768	55	US-60-230-435-751	Sequence 751, App	c 126	19	1.6	4417	28	US-09-710-286-3349	Sequence 3349, Ap
54	20	1.7	32768	55	US-60-234-446-423	Sequence 423, App	c 127	19	1.6	4417	28	US-09-717-350-4428	Sequence 4428, Ap
55	20	1.6	202	13	US-08-917-045-2919	Sequence 2919, Ap	c 128	19	1.6	4417	29	US-09-726-171-2420	Sequence 2420, Ap
56	19	1.6	202	21	US-09-540-229-91915	Sequence 91915, A	c 129	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
57	19	1.6	202	34	US-60-028-710-1175	Sequence 1175, Ap	c 130	19	1.6	4417	29	US-09-726-789-4562	Sequence 4562, Ap
58	19	1.6	202	35	US-60-032-868-1175	Sequence 1175, Ap	c 131	19	1.6	4417	29	US-09-726-789-4562	Sequence 22012, A
59	19	1.6	207	8	US-08-435-761-839	Sequence 839, App	c 132	19	1.6	6420	51	US-60-173-464-22012	Sequence 31697, A
60	19	1.6	207	20	US-09-532-315-27587	Sequence 27587, A	c 133	19	1.6	6420	51	US-60-191-637-31697	Sequence 2160, A
61	19	1.6	252	19	US-09-521-640-20483	Sequence 20483, A	c 134	19	1.6	7050	24	US-09-620-392-1799	Sequence 1799, Ap
62	19	1.6	252	20	US-09-539-806-42459	Sequence 42459, A	c 135	19	1.6	7050	24	US-09-620-392-1799	Sequence 10737, A
63	19	1.6	252	46	US-60-140-769-59	Sequence 59, App1	c 136	19	1.6	23119	24	US-09-620-392-8254	Sequence 8254, Ap
64	19	1.6	253	16	US-09-566-620-3901	Sequence 3901, Ap	c 137	19	1.6	23119	24	US-09-702-134-13316	Sequence 13316, A
65	19	1.6	263	13	US-08-978-620-5877	Sequence 5877, Ap	c 138	19	1.6	26601	24	US-09-620-392-43660	Sequence 43660, A
66	19	1.6	263	20	US-09-532-315-27599	Sequence 27599, A	c 139	19	1.6	26601	24	US-09-702-134-9967	Sequence 9967, Ap
67	19	1.6	265	44	US-60-127-461-6141	Sequence 6141, Ap	c 140	19	1.6	147215	54	US-09-534-859-682	Sequence 549, App
68	19	1.6	282	11	US-08-766-606-116	Sequence 116, App	c 141	19	1.6	187943	55	US-60-226-176-1377	Sequence 1377, Ap
69	19	1.6	282	13	US-08-979-854A-116	Sequence 116, App	c 142	19	1.6	187943	55	US-60-233-468-1377	Sequence 1377, Ap
70	19	1.6	282	20	US-09-532-315-27577	Sequence 27577, A	c 143	19	1.6	255488	19	US-09-528-237A-1892	Sequence 1892, Ap
71	19	1.6	328	16	US-09-205-070-38258	Sequence 38258, A	c 144	19	1.6	153	91	US-09-565-240-53157	Sequence 53157, A
72	19	1.6	328	17	US-09-321-214-29884	Sequence 29884, A	c 145	18	1.6	242	14	US-09-172-519-4647	Sequence 4647, Ap
73	19	1.6	328	17	US-09-340-623-38258	Sequence 38258, A	c 146	18	1.6	242	14	US-09-016-884-2666	Sequence 2666, Ap
74	19	1.6	328	19	US-09-516-335-29884	Sequence 29884, A	c 147	18	1.6	242	35	US-60-037-043-2666	Sequence 2666, Ap
75	19	1.6	328	29	US-09-733-811A-29884	Sequence 29884, A	c 148	18	1.6	244	12	US-08-865-594-5697	Sequence 5697, Ap
76	19	1.6	328	29	US-09-733-811A-29884	Sequence 29884, A	c 149	18	1.6	244	12	US-09-540-229-21369	Sequence 21369, A
77	19	1.6	333	21	US-09-540-499-20491	Sequence 20491, A	c 150	18	1.6	250	19	US-09-521-640-59480	Sequence 59480, A
78	19	1.6	333	47	US-60-156-625-3194	Sequence 3194, Ap	c 151	18	1.6	253	16	US-09-263-191-22384	Sequence 22384, A
79	19	1.6	390	7	US-08-337-601A-4376	Sequence 4376, Ap	c 152	18	1.6	254	17	US-09-304-517A-161300	Sequence 161300, A
80	19	1.6	390	9	US-08-530-112A-4376	Sequence 4376, Ap	c 153	18	1.6	254	17	US-09-304-517A-161300	Sequence 20065, A
81	19	1.6	397	16	US-09-287-618-12792	Sequence 12792, A	c 154	18	1.6	258	12	US-60-202-214-20065	Sequence 607, App
82	19	1.6	399	17	US-09-362-510A-8270	Sequence 8270, Ap	c 155	18	1.6	258	33	US-60-017-768-607	Sequence 19749, A
83	19	1.6	408	25	US-09-654-617-255046	Sequence 255046, Ap	c 156	18	1.6	259	17	US-09-306-349-19749	Sequence 4082, Ap
84	19	1.6	408	27	US-09-684-016-255046	Sequence 255046, Ap	c 157	18	1.6	259	49	US-60-172-362-4082	Sequence 92952, A
85	19	1.6	409	17	US-09-359-067-11362	Sequence 11362, A	c 158	18	1.6	262	25	US-09-654-617-92952	Sequence 92952, A
86	19	1.6	412	17	US-09-304-517A-214782	Sequence 214782, A	c 159	18	1.6	262	27	US-09-684-016-92952	Sequence 2730, Ap
87	19	1.6	412	22	US-09-552-086-10661	Sequence 10661, A	c 160	18	1.6	264	13	US-09-244-000A-2730	Sequence 5040, Ap
88	19	1.6	413	22	US-09-359-067-11363	Sequence 11363, A	c 161	18	1.6	273	13	US-08-951-198-5040	Sequence 212, App
89	19	1.6	413	17	US-09-359-067-11363	Sequence 11363, A	c 162	18	1.6	273	8	US-08-487-829-212	Sequence 67810, A
90	19	1.6	429	16	US-60-096-409-13031	Sequence 13031, A	c 163	18	1.6	297	19	US-09-521-620-67810	Sequence 98401, A
91	19	1.6	429	41	US-60-096-409-13031	Sequence 13031, A	c 164	18	1.6	302	21	US-09-540-229-98401	Sequence 1112, Ap
92	19	1.6	431	18	US-09-431-517-18493	Sequence 18493, A	c 165	18	1.6	311	12	US-08-839-968-1112	Sequence 1112, Ap
93	19	1.6	432	18	US-09-431-517-20292	Sequence 20292, A	c 166	18	1.6	311	13	US-08-951-198-1112	Sequence 1421, Ap
94	19	1.6	463	27	US-09-699-997-3587	Sequence 3587, Ap	c 167	18	1.6	315	13	US-09-107-532-1421	Sequence 1293, Ap
95	19	1.6	476	25	US-09-644-867-1361	Sequence 1361, Ap	c 168	18	1.6	322	31	US-09-803-719-1316	Sequence 1316, Ap
96	19	1.6	479	28	US-09-705-926-18345	Sequence 18345, Ap	c 169	18	1.6	322	31	US-09-803-719-1316	Sequence 3725, Ap
97	19	1.6	480	22	US-09-565-309A-40609	Sequence 40609, A	c 170	18	1.6	348	17	US-09-332-782-3725	Sequence 3725, Ap
98	19	1.6	480	22	US-09-565-309A-45465	Sequence 45465, A	c 171	18	1.6	348	17	US-09-515-694-3725	Sequence 3725, Ap
99	19	1.6	512	20	US-09-532-315-22122	Sequence 22122, A	c 172	18	1.6	353	52	US-60-202-213-5168	Sequence 5168, Ap
100	19	1.6	555	20	US-09-532-315-22110	Sequence 22110, A	c 173	18	1.6	374	5	US-08-104-507A-4386	Sequence 4386, Ap
101	19	1.6	643	28	US-09-705-926-10646	Sequence 10646, A	c 174	18	1.6	374	5	US-08-104-507C-4386	Sequence 4386, Ap
102	19	1.6	774	46	US-08-471-275-3739	Sequence 3739, Ap	c 175	18	1.6	393	16	US-09-293-972-26573	Sequence 26573, A
103	19	1.6	953	46	US-60-140-806-242	Sequence 242, App	c 176	18	1.6	395	19	US-09-521-640-95017	Sequence 95017, A
104	19	1.6	1179	16	US-09-248-796-5840	Sequence 5840, Ap	c 177	18	1.6	405	19	US-09-521-640-96181	Sequence 96181, A
105	19	1.6	1368	15	US-09-107-532-2819	Sequence 2819, Ap	c 178	18	1.6	419	16	US-09-274-861-2436	Sequence 2436, Ap
106	19	1.6	1455	19	US-09-505-532-11797	Sequence 11797, A	c 179	18	1.6	422	23	US-09-606-977-46838	Sequence 46838, Ap
107	19	1.6	1599	15	US-09-107-532-2530	Sequence 2530, Ap	c 180	18	1.6	442	23	US-09-306-350A-15245	Sequence 15245, A
108	19	1.6	1684	24	US-09-620-392-58938	Sequence 58938, A	c 181	18	1.6	442	17	US-09-654-617-24389	Sequence 24389, A
109	19	1.6	1832	27	US-09-684-016-16885	Sequence 16885, A	c 182	18	1.6	447	25	US-09-684-016-24389	Sequence 78957, A
110	19	1.6	1832	25	US-09-684-016-16885	Sequence 16885, A	c 183	18	1.6	456	23	US-09-606-977-78957	Sequence 1494, Ap
111	19	1.6	2572	25	US-09-644-867-5241	Sequence 5241, Ap	c 184	18	1.6	463	22	US-09-496-914A-1494	Sequence 13545, A
112	19	1.6	2572	25	US-09-652-814-8953	Sequence 8953, Ap	c 185	18	1.6	463	22	US-09-552-086-13545	Sequence 4081, Ap
113	19	1.6	3506	30	US-09-770-174-3978	Sequence 3978, Ap	c 186	18	1.6	482	23	US-09-606-977-69202	Sequence 69202, A
114	19	1.6	4120	51	US-60-173-464-22013	Sequence 22013, A	c 187	18	1.6	496	23	US-09-620-608-310	Sequence 310, App
115	19	1.6	4120	51	US-60-191-637-31698	Sequence 31698, A	c 188	18	1.6	500	24	US-09-620-608-310	Sequence 310, App
116	19	1.6	4121	51	US-60-191-637-31698	Sequence 31698, A	c 189	18	1.6	520	25	US-60-144-883-310	Sequence 452933, A
117	19	1.6	4121	51	US-60-191-637-31698	Sequence 25161, A	c 190	18	1.6	520	25	US-09-654-617-452933	Sequence 452933, A
118	19	1.6	4417	25	US-09-644-867-7077	Sequence 7077, Ap	c 191	18	1.6	520	25	US-09-684-016-452933	Sequence 41883, A
119	19	1.6	4417	25	US-09-644-867-7077	Sequence 6390, Ap	c 192	18	1.6	547	19	US-09-505-532-41883	Sequence 41883, A
120	19	1.6	4417	25	US-09-649-166-6390	Sequence 6390, Ap	c 193	18	1.6	547	19	US-09-505-532-41883	Sequence 41883, A

c 194	18	1.6	568	28	US-09-705-926-7311	Sequence 7311, Ap	267	18	1.6	14280	24	US-09-620-392-10818	Sequence 10818, A
c 195	18	1.6	585	44	US-60-125-818-421	Sequence 421, App	268	18	1.6	14280	24	US-09-702-134-16692	Sequence 16692, A
c 196	18	1.6	593	52	US-60-207-458-6733	Sequence 6733, Ap	c 269	18	1.6	16042	24	US-09-620-392-12451	Sequence 12451, A
c 197	18	1.6	598	17	US-09-354-899-8009	Sequence 8009, Ap	c 270	18	1.6	16042	24	US-09-702-134-26835	Sequence 26835, A
c 198	18	1.6	607	17	US-09-354-899-4447	Sequence 4447, Ap	c 271	18	1.6	17716	35	US-60-038-081-375	Sequence 375, App
c 199	18	1.6	627	49	US-60-172-519-4646	Sequence 4646, Ap	c 272	18	1.6	17716	35	US-60-046-933-385	Sequence 385, App
c 200	18	1.6	692	54	US-60-426-326-4152	Sequence 4152, Ap	c 273	18	1.6	17716	35	US-60-068-228-388	Sequence 388, App
c 201	18	1.6	714	17	US-09-399-720-12272	Sequence 12272, A	c 274	18	1.6	17716	35	US-09-702-134-2311	Sequence 2311, App
c 202	18	1.6	715	17	US-09-399-720-16743	Sequence 16743, A	c 275	18	1.6	18966	28	US-09-620-392-25724	Sequence 25724, A
c 203	18	1.6	738	24	US-09-620-608-1354	Sequence 1354, Ap	c 276	18	1.6	18966	28	US-60-213-847-152	Sequence 152, App
c 204	18	1.6	738	46	US-60-144-883-1354	Sequence 1354, Ap	c 277	18	1.6	20216	53	US-60-207-583-168	Sequence 168, App
c 205	18	1.6	784	11	US-08-781-883-1354	Sequence 537, App	c 278	18	1.6	20216	53	PCT-US01-01354-31985	Sequence 31985, A
c 206	18	1.6	784	13	US-08-956-171-537	Sequence 537, App	c 279	18	1.6	23394	52	US-60-207-583-283	Sequence 283, App
c 207	18	1.6	784	13	US-08-956-171-537	Sequence 729, App	c 280	18	1.6	32768	52	US-60-212-656-4	Sequence 4, Appl1
c 208	18	1.6	899	55	US-60-236-804-729	Sequence 12278, A	c 281	18	1.6	32768	55	US-60-230-435-382	Sequence 382, App
c 209	18	1.6	1131	50	US-60-182-316-12278	Sequence 1905, Ap	c 282	18	1.6	32768	55	US-60-230-445-992	Sequence 992, App
c 210	18	1.6	1134	25	US-09-654-617-1905	Sequence 1905, Ap	c 283	18	1.6	38023	12	US-08-831-156A-87	Sequence 87, Appl
c 211	18	1.6	1134	27	US-09-684-016-1905	Sequence 1905, Ap	c 284	18	1.6	38023	12	US-08-831-156A-88	Sequence 88, Appl
c 212	18	1.6	1260	1	PCT-US00-08016-1	Sequence 1, Appl1	c 285	18	1.6	38522	12	US-60-167-324-1488	Sequence 1488, Ap
c 213	18	1.6	1260	16	US-09-285-516-1	Sequence 1, Appl1	c 286	18	1.6	40657	48	US-60-173-386-1430	Sequence 1430, Ap
c 214	18	1.6	1266	45	US-60-138-103-25058	Sequence 25058, A	c 287	18	1.6	40657	48	US-60-173-386-1430	Sequence 1470, Ap
c 215	18	1.6	1281	12	US-08-827-356-561	Sequence 561, App	c 288	18	1.6	40958	49	US-60-175-871-1610	Sequence 1764, A
c 216	18	1.6	1281	14	US-09-035-744-1	Sequence 1, Appl1	c 289	18	1.6	40958	49	US-60-184-775-1470	Sequence 1576, Ap
c 217	18	1.6	1281	23	US-09-611-529-2482	Sequence 2482, Ap	c 290	18	1.6	40958	50	US-60-191-637-41764	Sequence 429, App
c 218	18	1.6	1381	1	PCT-US01-01347-177	Sequence 177, App	c 291	18	1.6	40962	51	US-60-258-272-14	Sequence 57, Appl
c 219	18	1.6	1392	11	US-08-785-047-1	Sequence 1, Appl1	c 292	18	1.6	41094	24	US-60-245-228-30	Sequence 30, Appl
c 220	18	1.6	1392	11	US-08-785-047A-1	Sequence 1, Appl1	c 293	18	1.6	45921	56	US-60-245-228-30	Sequence 14, Appl
c 221	18	1.6	1392	13	US-08-913-586-1	Sequence 1, Appl1	c 294	18	1.6	45921	56	US-60-245-228-30	Sequence 15, Appl
c 222	18	1.6	1404	12	US-08-827-356-930	Sequence 930, App	c 295	18	1.6	45921	56	US-60-245-228-30	Sequence 685, App
c 223	18	1.6	1404	23	US-09-611-529-825	Sequence 825, App	c 296	18	1.6	61010	56	US-60-233-166-178650	Sequence 1055, Ap
c 224	18	1.6	1673	23	US-09-471-275-7068	Sequence 7068, Ap	c 297	18	1.6	61010	56	US-60-233-166-178650	Sequence 142, App
c 225	18	1.6	1679	23	US-09-607-200-6698	Sequence 6698, Ap	c 298	18	1.6	61010	56	US-60-233-166-178650	Sequence 429, App
c 226	18	1.6	1745	18	US-09-457-877-112	Sequence 112, App	c 299	18	1.6	79375	20	US-09-534-859-574	Sequence 574, App
c 227	18	1.6	1843	28	US-09-702-134-19051	Sequence 19051, A	c 300	18	1.6	79375	20	US-09-534-859-574	Sequence 142, App
c 228	18	1.6	1908	24	US-09-620-392-62559	Sequence 62559, A	c 301	18	1.6	90425	20	US-09-534-859-574	Sequence 142, App
c 229	18	1.6	2114	44	US-60-126-248-532	Sequence 532, App	c 302	18	1.6	90425	20	US-09-534-859-574	Sequence 142, App
c 230	18	1.6	2114	44	US-60-126-248-532	Sequence 532, App	c 303	18	1.6	90425	20	US-09-534-859-574	Sequence 142, App
c 231	18	1.6	2114	49	US-60-172-373-21742	Sequence 21742, A	c 304	18	1.6	90425	20	US-09-534-859-574	Sequence 142, App
c 232	18	1.6	2162	18	US-09-404-520-12569	Sequence 12569, A	c 305	18	1.6	100815	27	US-09-534-859-574	Sequence 142, App
c 233	18	1.6	3212	17	US-09-338-425-2523	Sequence 2523, Ap	c 306	18	1.6	103388	19	US-09-528-237A-1041	Sequence 82, Appl
c 234	18	1.6	3212	25	US-09-644-868-10046	Sequence 10046, A	c 307	18	1.6	103388	19	US-09-528-237A-1041	Sequence 1041, Ap
c 235	18	1.6	3212	29	US-09-698-010-15328	Sequence 15328, A	c 308	18	1.6	103388	19	US-09-528-237A-1041	Sequence 7448, Ap
c 236	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 309	18	1.6	122075	58	US-60-261-976-31	Sequence 31, Appl
c 237	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 310	18	1.6	122075	58	US-60-261-976-31	Sequence 149, App
c 238	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 311	18	1.6	122075	58	US-60-261-976-31	Sequence 1556, Ap
c 239	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 312	18	1.6	122075	58	US-60-261-976-31	Sequence 1438, Ap
c 240	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 313	18	1.6	122075	58	US-60-261-976-31	Sequence 178650, A
c 241	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 314	18	1.6	122075	58	US-60-261-976-31	Sequence 14333, A
c 242	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 315	18	1.6	122075	58	US-60-261-976-31	Sequence 14275, A
c 243	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 316	18	1.6	122075	58	US-60-261-976-31	Sequence 1649, Ap
c 244	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 317	18	1.6	122075	58	US-60-261-976-31	Sequence 7928, Ap
c 245	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 318	18	1.6	122075	58	US-60-261-976-31	Sequence 44609, A
c 246	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 319	18	1.6	122075	58	US-60-261-976-31	Sequence 4252, Ap
c 247	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 320	18	1.6	122075	58	US-60-261-976-31	Sequence 4252, Ap
c 248	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 321	18	1.6	122075	58	US-60-261-976-31	Sequence 1743, Ap
c 249	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 322	18	1.6	122075	58	US-60-261-976-31	Sequence 486, App
c 250	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 323	18	1.6	122075	58	US-60-261-976-31	Sequence 1568, Ap
c 251	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 324	18	1.6	122075	58	US-60-261-976-31	Sequence 14739, A
c 252	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 325	18	1.6	122075	58	US-60-261-976-31	Sequence 14739, A
c 253	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 326	18	1.6	122075	58	US-60-261-976-31	Sequence 49371, A
c 254	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 327	18	1.6	122075	58	US-60-261-976-31	Sequence 90929, A
c 255	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 328	18	1.6	122075	58	US-60-261-976-31	Sequence 83651, A
c 256	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 329	18	1.6	122075	58	US-60-261-976-31	Sequence 83651, A
c 257	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 330	18	1.6	122075	58	US-60-261-976-31	Sequence 197604, A
c 258	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 331	18	1.6	122075	58	US-60-261-976-31	Sequence 6160, Ap
c 259	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 332	18	1.6	122075	58	US-60-261-976-31	Sequence 3999, A
c 260	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 333	18	1.6	122075	58	US-60-261-976-31	Sequence 22838, A
c 261	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 334	18	1.6	122075	58	US-60-261-976-31	Sequence 29116, A
c 262	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 335	18	1.6	122075	58	US-60-261-976-31	Sequence 14150, A
c 263	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 336	18	1.6	122075	58	US-60-261-976-31	Sequence 172, App
c 264	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 337	18	1.6	122075	58	US-60-261-976-31	Sequence 28271, A
c 265	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 338	18	1.6	122075	58	US-60-261-976-31	Sequence 172, App
c 266	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 339	18	1.6	122075	58	US-60-261-976-31	Sequence 14132, A
	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap		18	1.6	122075	58	US-09-304-517A-146422	Sequence 146422, A
	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap		18	1.6	122075	58	US-09-552-066-19293	Sequence 19293, A

us-09-151-409-15.oligo.rnmpm

Tue May 15 07:26:54 2001

340	17	1.5	239	17	US-09-304-517A-161586	Sequence 161586,	413	17	1.5	322	35	US-60-036-549-2160	Sequence 2160, Ap
341	17	1.5	240	17	US-09-304-517A-205485	Sequence 205485,	414	17	1.5	323	53	US-60-211-750-130772	Sequence 60772, A
342	17	1.5	241	17	US-09-263-191-23026	Sequence 23026, A	c 415	17	1.5	324	53	US-60-217-080-13030	Sequence 13302, A
343	17	1.5	242	13	US-08-917-045-5313	Sequence 5313, Ap	c 416	17	1.5	331	17	US-09-394-745-34427	Sequence 34427, A
344	17	1.5	242	17	US-09-304-517A-153598	Sequence 153598,	c 417	17	1.5	331	19	US-09-521-640-96057	Sequence 96057, A
345	17	1.5	242	17	US-09-534-843-11142	Sequence 11142, A	c 418	17	1.5	331	22	US-09-563-306-54992	Sequence 54992, A
346	17	1.5	242	34	US-60-028-710-3369	Sequence 3369, Ap	c 419	17	1.5	331	25	US-09-684-016-326058	Sequence 326058,
347	17	1.5	242	35	US-60-032-838-3569	Sequence 3569, Ap	c 420	17	1.5	331	25	US-09-684-016-326058	Sequence 326058,
348	17	1.5	247	16	US-09-081-109-1377	Sequence 1377, Ap	c 421	17	1.5	339	16	US-09-283-466-2023	Sequence 2023, Ap
349	17	1.5	247	16	US-09-244-000A-57733	Sequence 57733, A	c 422	17	1.5	341	53	US-09-304-517A-281337	Sequence 281337,
350	17	1.5	247	16	US-09-263-191-30715	Sequence 30715, A	c 423	17	1.5	341	53	US-60-211-750-13492	Sequence 33492, A
351	17	1.5	247	16	US-09-283-466-39490	Sequence 39490, A	c 424	17	1.5	343	16	US-09-270-767-12437	Sequence 12437, A
352	17	1.5	247	21	US-09-540-764-43006	Sequence 43006, A	c 425	17	1.5	349	18	US-09-304-517A-276368	Sequence 276368,
353	17	1.5	247	36	US-09-048-727-1377	Sequence 1377, Ap	c 426	17	1.5	349	18	US-09-440-687-29614	Sequence 29614, A
354	17	1.5	248	17	US-09-304-517A-200417	Sequence 200417,	c 427	17	1.5	352	11	US-60-144-084-17177	Sequence 17177, A
355	17	1.5	248	17	US-09-304-517A-146411	Sequence 146411,	c 428	17	1.5	352	13	US-08-766-606-370	Sequence 370, App
356	17	1.5	249	22	US-09-552-086-19319	Sequence 19319, A	c 429	17	1.5	352	20	US-08-979-854A-370	Sequence 27578, A
357	17	1.5	249	22	US-09-565-240-31174	Sequence 31174, A	c 430	17	1.5	352	20	US-09-532-315-27578	Sequence 40790, A
358	17	1.5	249	45	US-09-133-691-5297	Sequence 5297, Ap	c 431	17	1.5	353	53	US-09-211-750-40790	Sequence 395805,
359	17	1.5	250	21	US-09-304-517A-199449	Sequence 199449,	c 432	17	1.5	353	25	US-09-654-617-395805	Sequence 395805,
360	17	1.5	250	21	US-09-540-229-117439	Sequence 117439,	c 433	17	1.5	356	22	US-09-684-016-395805	Sequence 29883, A
361	17	1.5	252	16	US-09-263-191-15479	Sequence 15479, A	c 434	17	1.5	361	22	US-09-565-240-29883	Sequence 3837, Ap
362	17	1.5	252	16	US-09-294-093A-3378	Sequence 3378, A	c 435	17	1.5	361	50	US-60-189-657-1987	Sequence 1987, Ap
363	17	1.5	252	16	US-09-284-093B-3849	Sequence 3849, Ap	c 436	17	1.5	361	45	US-60-133-691-3837	Sequence 675, App
364	17	1.5	252	40	US-09-082-567-3849	Sequence 3849, Ap	c 437	17	1.5	363	14	US-09-016-434-675	Sequence 11143, A
365	17	1.5	254	17	US-09-304-517A-158260	Sequence 158260,	c 438	17	1.5	363	20	US-09-534-843-11143	Sequence 8959, Ap
366	17	1.5	255	17	US-09-306-349-8390	Sequence 8390, Ap	c 439	17	1.5	364	18	US-09-431-517-8959	Sequence 93842, A
367	17	1.5	258	16	US-09-244-000A-42533	Sequence 42533, A	c 440	17	1.5	365	19	US-09-521-640-93842	Sequence 42177, A
368	17	1.5	260	13	US-08-917-045-2658	Sequence 2658, Ap	c 441	17	1.5	367	22	US-09-565-240-42177	Sequence 10471, A
369	17	1.5	260	21	US-09-540-229-73215	Sequence 73215, A	c 442	17	1.5	367	23	US-09-606-977-10471	Sequence 3542, Ap
370	17	1.5	260	34	US-60-028-710-914	Sequence 914, App	c 443	17	1.5	367	45	US-09-515-128-19652	Sequence 19652, A
371	17	1.5	261	17	US-09-032-838-914	Sequence 914, App	c 444	17	1.5	368	19	US-09-515-128-19652	Sequence 30194, A
372	17	1.5	261	17	US-09-304-517A-189371	Sequence 189371,	c 445	17	1.5	368	22	US-09-565-240-30194	Sequence 8337, Ap
373	17	1.5	261	21	US-09-540-229-181527	Sequence 181527,	c 446	17	1.5	368	45	US-09-577-408-8337	Sequence 4210, Ap
374	17	1.5	261	40	US-09-089-037-425	Sequence 425, App	c 447	17	1.5	371	26	US-60-133-691-4210	Sequence 2700, Ap
375	17	1.5	262	11	US-08-733-814-275	Sequence 275, App	c 448	17	1.5	371	26	US-09-663-779-2700	Sequence 11402, A
376	17	1.5	262	17	US-09-304-517A-183720	Sequence 183720,	c 449	17	1.5	373	52	US-09-270-767-11402	Sequence 51800, A
377	17	1.5	262	17	US-09-333-535A-9682	Sequence 9682, Ap	c 450	17	1.5	373	53	US-60-207-458-51800	Sequence 13492, A
378	17	1.5	262	21	US-09-540-499-25360	Sequence 25360, A	c 451	17	1.5	375	53	US-60-217-080-13492	Sequence 97830, A
379	17	1.5	262	21	US-09-540-764-27327	Sequence 27327, A	c 452	17	1.5	376	19	US-09-521-640-97830	Sequence 193783, A
380	17	1.5	262	22	US-09-552-086-23067	Sequence 23067, A	c 453	17	1.5	378	19	US-09-521-640-193783	Sequence 20641, A
381	17	1.5	262	42	US-60-108-390-125	Sequence 125, App	c 454	17	1.5	378	46	US-60-140-769-20641	Sequence 6146, Ap
382	17	1.5	265	16	US-09-263-191-25425	Sequence 25425, A	c 455	17	1.5	380	20	US-60-202-213-6146	Sequence 16365, A
383	17	1.5	265	17	US-09-304-517A-156611	Sequence 156611,	c 456	17	1.5	380	25	US-09-534-843-16366	Sequence 459086,
384	17	1.5	265	20	US-09-539-806-47440	Sequence 47440, A	c 457	17	1.5	384	25	US-09-654-617-459086	Sequence 459086,
385	17	1.5	265	20	US-09-244-000A-5512	Sequence 5512, Ap	c 458	17	1.5	384	23	US-09-606-977-8301	Sequence 8301, Ap
386	17	1.5	266	16	US-09-304-517A-285706	Sequence 285706,	c 459	17	1.5	387	19	US-09-505-533-22153	Sequence 2153, A
387	17	1.5	266	17	US-09-199-129A-4159	Sequence 4159, Ap	c 460	17	1.5	387	19	US-09-521-640-8764	Sequence 8764, Ap
388	17	1.5	272	15	US-09-304-517A-278709	Sequence 278709,	c 461	17	1.5	387	19	US-09-362-510-23801	Sequence 23801, A
389	17	1.5	272	17	US-09-283-466-34185	Sequence 34185, A	c 462	17	1.5	391	17	US-09-362-510A-23801	Sequence 23801, A
390	17	1.5	273	16	US-09-244-000A-5512	Sequence 5512, Ap	c 463	17	1.5	391	17	US-09-521-640-298557	Sequence 298557, A
391	17	1.5	273	52	US-60-209-830-42825	Sequence 42825, A	c 464	17	1.5	391	17	US-09-521-640-176172	Sequence 176172, A
392	17	1.5	274	16	US-09-244-000A-63998	Sequence 63998, A	c 465	17	1.5	393	51	US-60-197-873-40964	Sequence 40964, A
393	17	1.5	274	17	US-09-304-517A-294632	Sequence 294632,	c 466	17	1.5	397	6	US-08-220-691-2862	Sequence 2862, Ap
394	17	1.5	274	22	US-09-552-086-4321	Sequence 4321, Ap	c 467	17	1.5	397	6	US-08-220-691B-2862	Sequence 2862, Ap
395	17	1.5	276	16	US-09-199-129A-2388	Sequence 2388, Ap	c 468	17	1.5	397	23	US-09-606-977-26342	Sequence 26342, A
396	17	1.5	276	16	US-09-215-179A-4949	Sequence 4949, Ap	c 469	17	1.5	399	19	US-09-528-409-67394	Sequence 67394, A
397	17	1.5	278	16	US-09-244-000A-98249	Sequence 98249, A	c 470	17	1.5	400	18	US-09-421-106-23844	Sequence 23844, A
398	17	1.5	281	52	US-60-207-458-21127	Sequence 21127, A	c 471	17	1.5	400	19	US-09-521-640-298557	Sequence 298557, A
399	17	1.5	282	15	US-09-159-129A-1731	Sequence 1731, Ap	c 472	17	1.5	401	24	US-09-637-086A-38051	Sequence 38051, A
400	17	1.5	282	21	US-09-540-499-20865	Sequence 20865, A	c 473	17	1.5	405	16	US-09-248-797-3903	Sequence 3903, Ap
401	17	1.5	283	11	US-08-798-074-7560	Sequence 7560, Ap	c 474	17	1.5	405	17	US-09-346-956-76	Sequence 76, Appl
402	17	1.5	285	16	US-09-244-000A-19723	Sequence 19723, A	c 475	17	1.5	405	22	US-09-565-240-53016	Sequence 53016, A
403	17	1.5	285	17	US-09-304-517A-289315	Sequence 289315,	c 476	17	1.5	408	22	US-09-528-409-82166	Sequence 82166, A
404	17	1.5	287	17	US-09-306-349-39619	Sequence 39619, A	c 477	17	1.5	408	45	US-09-565-240-29718	Sequence 29718, A
405	17	1.5	288	17	US-09-304-517A-188395	Sequence 188395,	c 478	17	1.5	408	45	US-09-133-691-10963	Sequence 10963, A
406	17	1.5	291	20	US-09-534-843-16363	Sequence 16363, A	c 479	17	1.5	409	22	US-09-565-306-29369	Sequence 29369, A
407	17	1.5	306	14	US-09-070-694-3294	Sequence 3294, Ap	c 480	17	1.5	409	25	US-09-654-617-292009	Sequence 292009,
408	17	1.5	317	25	US-09-654-617-26497	Sequence 26497, A	c 481	17	1.5	409	27	US-09-684-016-292009	Sequence 292009,
409	17	1.5	317	27	US-09-684-016-26497	Sequence 26497, A	c 482	17	1.5	410	19	US-09-521-640-228973	Sequence 228973,
410	17	1.5	319	21	US-09-540-499-4660	Sequence 4660, Ap	c 483	17	1.5	411	50	US-60-189-657-35493	Sequence 3549, Ap
411	17	1.5	322	14	US-09-036-591-2160	Sequence 2160, Ap	c 484	17	1.5	411	50	US-60-207-458-33443	Sequence 33443, A
412	17	1.5					c 485	17	1.5	416	52		


```
QY 481 ccatggagagtcggttatactagccatagccagagataagctaaaaaaattattgac 540
Db 481 ccatggagagtcggttatactagccatagccatagccagagataagctaaaaaaattattgac 540
QY 541 gatcttgacgcaaaagaaacatgaattacaaacacagaatgagaagtattctctgcagaac 600
Db 541 gatcttgacgcaaaagaaacatgaattacaaacacagaatgagaagtattctctgcagaac 600
QY 601 ggtgatggttaactctctaggaagttatagaagattcttgacgcaaaacaaatcccgcaatacaa 660
Db 601 ggtgatggttaactctctaggaagttatagaagattcttgacgcaaaacaaatcccgcaatacaa 660
QY 661 aatatagctttacgttcacgaaacaaagagacttaaaagcgagattagagaatgcaatggaa 720
Db 661 aatatagctttacgttcacgaaacaaagagacttaaaagcgagattagagaatgcaatggaa 720
QY 721 gttcaggaagagatttttaagagagctggtaaccttggttagatcaggtttacacaattatat 780
Db 721 gttcaggaagagatttttaagagagctggtaaccttggttagatcaggtttacacaattatat 780
QY 781 actaaacataatagtaattaccacaacataatgcacaaagctggcagacttgacctgaga 840
Db 781 actaaacataatagtaattaccacaacataatgcacaaagctggcagacttgacctgaga 840
QY 841 caaaaggctgaatatctaaaaagccttaataatggtggtgagagctgttacaagagtta 900
Db 841 caaaaggctgaatatctaaaaagccttaataatggtggtgagagctgttacaagagtta 900
QY 901 aatatcgatgctgcgactagctctcagacagatactctggaagaaagtcacagaacgtgct 960
Db 901 aatatcgatgctgcgactagctctcagacagatactctggaagaaagtcacagaacgtgct 960
QY 961 gacaagtttgagatagaacaaatacgttlaaaacttaagaatagtgacttaagtttaatt 1020
Db 961 gacaagtttgagatagaacaaatacgttlaaaacttaagaatagtgacttaagtttaatt 1020
QY 1021 aataaagcgttaaaagatcataatgatgatggttaactgaaggttgagtaagttaagtaag 1080
Db 1021 aataaagcgttaaaagatcataatgatgatggttaactgaaggttgagtaagttaagtaag 1080
QY 1081 aaactacgttaaaatgataaaacactactctgaaaaagctagtaaaattcaagaattagag 1140
Db 1081 aaactacgttaaaatgataaaacactactctgaaaaagctagtaaaattcaagaattagag 1140
QY 1141 gcacgttaagtaaaagcgtt 1158
Db 1141 gcacgttaagtaaaagcgtt 1158
```

RESULT 2

```
US-07-945-954A-2
; Sequence 2, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
```

```
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
; US-07-945-954A-2
```

```
Query Match 34.7%; Score 402; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 8.9e-197;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 190 ggtccgcgcgtgactaggggtacaataaatgacccgcaaaagacaaaagaagctctttgac 249
Db 340 GGATCCGCGGTGACTAGGGGTACAATAATGACCCCAAGAGCAAAAGAGCTCTTGAC 399
QY 250 aagtatgagctagaacacccatgacttcaaaactaagaatgaagggtttaaactgagaat 309
Db 400 AAGTATGAGCTAGAAAACCATGACTTTAAAACTAAGAATGAAGGTTAAAACTGAGAA 459
QY 310 gaagggtttaaactgagaatgaagggtttaaactgagaatgaagggtttaaactgag 369
Db 460 GAAGGGTTAAACTGAGAACTGAAGGTTAAAACCTGAGAACTGAAGGTTAAAACTGAG 519
QY 370 gtgcagagagtggttctctagggggcggtagaaaaacccgggacaaagcacgagaactctt 429
Db 520 GTCGACAGAGTGTTCCTCTAGGGGACGCTAGAAAACCCGGACAAAGCAGACGAACTTCT 579
QY 430 aacaagtatgacgtagagaactctatgtttacaagctaatatgacaagtaccatggaga 489
Db 580 ACAAGTATGACGTAGAGAACTCTATGTTACAAAGCTAATAATGACAAGTTACCATGGAG 639
QY 490 gtgcgttatactagcgacacgcagagaataagctaaaaaaattattgacgactcttgc 591
Db 640 GTGCGTTATACTAGGCATAGCCGACAGAGTAAGCTAAAAAAATTTATTGACGATCTTGAC 699
QY 550 gcaaaagaacatgaattacaacaacagaaatgagaagttatct 591
Db 700 GCAAAAGAACAATGAATTACAACAACAGAAATGAGAAGTTATCT 741
```

RESULT 3

```
US-07-945-954A-1
; Sequence 1, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
```

```

;
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..861
; US-07-945-954A-1

Query Match 20.7%; Score 240; DB 3; Length 861;
Best Local Similarity 100.0%; Pred. No. 5.6e-113; Indels 0; Gaps 0;
Matches 240; Conservative 0; Mismatches 0;

QY 910 gtgcgactaggtctcagacagatactctggaagaaagtcacagaacgtgctgacaagttt 969
Db 4 GTCCGCACTAGGTCTCAGACACATACTCTGGAAAAGTACAAAGACGTGCTGACAAGTTT 63

QY 970 gagatgaaaaacaatacgtttaaaacttaagaatagtgacttaagttttaataataaagcg 1029
Db 64 GAGATAGAAAACAATAACGTTAAAGACTTAAGATACTGAAGTAAATTAATAATAAAGCG 123

QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGAATCATATGATGAGTTTAAGTGAAGAGTTGAGTAATGCTTAAGAGAAACTACGT 183

QY 1090 aaaaatgataaacactatctgaaaagctagtaaaattcaagaattagaggcagtaag 1149
Db 184 AAAATGATAATCACTATCTGAAAAGCTAGTAAATTCAGAATTTAGAGGCACGTAAG 243

RESULT 4
US-07-945-954A-7
; Sequence 7, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..519
; US-07-945-954A-7

Query Match 18.7%; Score 217; DB 3; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.4e-101; Indels 0; Gaps 0;
Matches 217; Conservative 0; Mismatches 0;

QY 910 gtgcgactaggtctcagacagatactctggaagaaagtcacagaacgtgctgacaagttt 969
Db 4 GTCCGCACTAGGTCTCAGACACATACTCTGGAAAAGTACAAAGACGTGCTGACAAGTTT 63

QY 970 gagatgaaaaacaatacgtttaaaacttaagaatagtgacttaagttttaataataaagcg 1029
Db 64 GAGATAGAAAACAATAACGTTAAAGACTTAAGATACTGAAGTAAATTAATAATAAAGCG 123

QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGAATCATATGATGAGTTTAAGTGAAGAGTTGAGTAATGCTTAAGAGAAACTACGT 183

QY 1090 aaaaatgataaacactatctgaaaagctagtaaaattcaagaattagaggcagtaag 1126
Db 184 AAAATGATAATCACTATCTGAAAAGCTAGTAAATTCAGAATTTAGAGGCACGTAAG 220

RESULT 5
US-07-945-954A-6
; Sequence 6, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/945,954A
;; FILING DATE: 16-SEP-1992
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 372.5672P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 741 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Streptococcus pyogenes
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..741
;; US-07-945-954A-6

Query Match 18.7%; Score 217; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaaagtagcaaacgctgctgacaagttt 969
|||||
Db 406 GTCGCGACTAGGTCTCAGACAGATACTCTGGAAGAAAGTACAAAGACGTGCTGACAAGTTT 465
|||||
QY 970 gagatagaaaaacaatacgttaaaacttaagaatactagtagtacttaagtttttaataaagcg 1029
|||||
Db 466 GAGATAGAAAACAACATACGTTAAACCTTAAGAATAGTAGTAAAGTTTATATAAAGCG 525
|||||
QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaacttaaaagagaactacgt 1089
|||||
Db 526 TTAAGAGATCATATGATGAGTTAACTGAAGAGTTTGAGTAATCTAAAGAGAACTACGT 585
|||||
QY 1090 aaaaatgataaatacactatctgaaaaagctagtaaaa 1126
|||||
Db 586 AAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAA 622
|||||

RESULT 6
US-07-945-954A-3
; Sequence 3, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992

;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 372.5672P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 822 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Streptococcus pyogenes
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..822
;; US-07-945-954A-3

Query Match 18.7%; Score 217; DB 3; Length 822;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaaagtagcaaacgctgctgacaagttt 969
|||||
Db 4 GTCGCGACTAGGTCTCAGACAGATACTCTGGAAGAAAGTACAAAGACGTGCTGACAAGTTT 63
|||||
QY 970 gagatagaaaaacaatacgttaaaacttaagaatactagtagtacttaagtttttaataaagcg 1029
|||||
Db 64 GAGATAGAAAACAACATACGTTAAACCTTAAGAATAGTAGTAAAGTTTATATAAAGCG 123
|||||
QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtagtaactgctaaagagaactacgt 1089
|||||
Db 124 TTAAGAGATCATATGATGAGTTAACTGAAGAGTTTGAGTAATCTGCTAAAGAGAACTACGT 183
|||||
QY 1090 aaaaatgataaatacactatctgaaaaagctagtaaaa 1126
|||||
Db 184 AAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAA 220
|||||

RESULT 7
US-07-945-954A-8
; Sequence 8, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.

```
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1029
US-07-945-954A-8

Query Match          18.7%; Score 217; DB 3; Length 1029;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaagaaagtacaaagacgtgctgacaagttt 969
DB 406 gtccgcgactaggtctcagacagatactctggaagaaagtacaaagacgtgctgacaagttt 465

QY 970 gagatagaaaaacaatacgtttaaacttaagaatagtgacttaagttaataataaagcg 1029
DB 466 GAGATAGAAAAACAATACGTTTAAACTTAAAGAAATAGTGACTTAACTTAAATAATAAAGCG 525

QY 1030 ttaaaagatcaataatgatgagtttaactgaagagtggtgagtaagtctaagagaaactcgt 1089
DB 526 TTAAGAATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 585

QY 1090 aaaaatgataaatcactatctgaaaaagctagtataaa 1126
DB 586 AAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAA 622

RESULT 8
US-08-325-278-5
; Sequence 5, Application US/08325278
; GENERAL INFORMATION:
; APPLICANT: Bivick, Lars
; APPLICANT: Sjöbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,278
; FILING DATE: 26-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMeesters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 450023.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4500
```

```
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1329
US-08-325-278-5

Query Match          13.0%; Score 151; DB 7; Length 1332;
Best Local Similarity 100.0%; Pred. No. 6.1e-67;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 acggtgatgttaactccttagggaagttatagaagatcttgcagcaacaatccccgaata 657
DB 1 AACCGTGATGGTAAATCCTAGGGAAGTTATAGAAGATCTTCCAGCAAAACAATCCCGCAATA 60

QY 658 caaaatatacgttttaacgttcacgaaaaaacaaggaccttaaaagcgagattagagaatgcaatg 717
DB 61 CAAAATATACGTTTACGTCACGAAAAACAAGGACTTAAAAAGCGAGATTAGAGAATGCAATG 120

QY 718 gaagttgcagggaagagatatttaagagagctg 748
DB 121 GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG 151

RESULT 9
US-07-945-954A-4
; Sequence 4, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```



```
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
US-07-945-954A-4

Query Match      4.6%; Score 53; DB 3; Length 561;
Best Local Similarity 100.0%; Pred. No. 3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacaataaataatgacccgcgaagaagcaaaagaagc 241
      |||||||
Db 139 GGATCCCGCGTGACTAGGGGTACAATAAATGACCCGCCAAGAGCAAAAGAAGC 191

RESULT 10
US-07-945-954A-9
; Sequence 9, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
US-07-945-954A-10

Query Match      4.5%; Score 52; DB 3; Length 408;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacaataaataatgacccgcgaagaagcaaaagaag 241
      |||||||
Db 49 GGATCCCGCGTGACTAGGGGTACAATAAATGACCCGCCAAGAGCAAAAGAAG 100

RESULT 12
US-07-945-954A-5
; Sequence 5, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia

Query Match      4.5%; Score 52; DB 3; Length 204;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacaataaataatgacccgcgaagaagcaaaagaag 241
      |||||||
Db 49 GGATCCCGCGTGACTAGGGGTACAATAAATGACCCGCCAAGAGCAAAAGAAG 100
```

```

STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,954A
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372,5672P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 918 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..915
US-07-945-954A-5

Query Match 4.5% Score 52; DB 3; Length 918;
Best Local Similarity 100.0%; Pred. No. 9,9e-16;
Matches 52; Conservative 0; Mismatches 0; Indels 0

QY 190 ggatccgcgcgtgactagggggtacataataatgaccgcgaagaagcaaaagaag 241
|||||
DB 49 GGATCCGCCGCTGACTAGGGGTACAATAATGACCCGCAAGCAAAAAGAAG 100
|||||

RESULT 13
US-07-945-860A-2
Sequence 2, Application US/07945860A
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: Antigen of Hybrid M Protein and Carrier
TITLE OF INVENTION: for Group A Streptococcal Vaccine
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,860A
FILING DATE: 19920916
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372,5770P

```

Query Match 4.1% Score 48; DB 13; Length 765;
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 atggtcgcgactaggctctcagacagatactctggaaaaagtacaagaa 54
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ATGGTCGCGACTAGGTCTCAGACAGATACTCTGGAAGTACAAGAA 48

RESULT 15
 US-08-914-479-3
 ; Sequence 3, Application US/08914479
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James B.
 ; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
 ; FILE REFERENCE: 48112.404C2
 ; CURRENT APPLICATION NUMBER: US/08/914.479
 ; CURRENT FILING DATE: 1997-08-19
 ; PRIOR APPLICATION NUMBER: 08/409,270
 ; PRIOR FILING DATE: 1995-03-23
 ; PRIOR APPLICATION NUMBER: 07/945,860
 ; PRIOR FILING DATE: 1992-09-16
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 765
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: An antigen of M5 and a carrier of the
 ; OTHER INFORMATION: COOH-terminal portion of M5
 US-08-914-479-3

Query Match 4.1% Score 48; DB 13; Length 765;
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 atggtcgcgactaggctctcagacagatactctggaaaaagtacaagaa 54
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 atggtcgcgactaggctctcagacagatactctggaaaaagtacaagaa 48

Search completed: May 13, 2001, 06:46:59
 Job time: 3538 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 05:29:41 ; Search time 86.88 Seconds
(without alignments)
5000.619 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatgctgcgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table:

OLIGO_NUC

GAP60 60.0 , Gapext 60.0

Searched: 262727 seqs, 187588001 residues

Word size : 10

Total number of hits satisfying chosen parameters: 122670

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending_Patents_NA_New.*

- 1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.6	275	5	US-09-540-212A-66707
2	18	1.6	784	4	US-08-956-171C-537
3	18	1.6	4813	1	PCT-US01-04098A-471
4	18	1.6	5623	1	PCT-US01-04098A-2439
5	18	1.6	283061	6	US-60-248-505-134
6	17	1.5	1332	5	US-09-540-212A-13409
7	17	1.5	1682	5	US-09-801-833-7074
8	17	1.5	3929	5	US-09-543-679A-2639
9	17	1.5	9834	4	US-08-956-171C-37
10	17	1.5	12447	5	US-09-543-679A-2640
11	17	1.5	126883	6	US-60-248-505-614
12	17	1.5	924430	5	US-09-335-032-12216
13	16	1.4	95	5	US-09-540-212A-37688
14	16	1.4	120	5	US-09-540-212A-21769
15	16	1.4	165	5	US-09-815-343-1168
16	16	1.4	240	5	US-09-815-343-220
17	16	1.4	245	5	US-09-540-212A-2062
18	16	1.4	274	5	US-09-540-212A-57406
19	16	1.4	266	6	US-60-010-803-3361
20	16	1.4	303	5	US-09-487-566A-3760
21	16	1.4	317	5	US-09-724-866A-19388
22	16	1.4	375	5	US-09-540-212A-61419
23	16	1.4	402	4	US-08-276-1630-5229
24	16	1.4	404	5	US-09-737-223-18462
25	16	1.4	426	1	PCT-US01-01339-992
26	16	1.4	426	1	PCT-US01-01339-992
27	16	1.4	452	5	US-09-737-223-9871

28	16	1.4	453	1	PCT-US01-01339-9895
29	16	1.4	473	5	US-09-737-223-31535
30	16	1.4	474	5	US-09-737-223-16219
31	16	1.4	486	5	US-09-724-866A-6168
32	16	1.4	489	5	US-09-724-866A-5608
33	16	1.4	506	5	US-09-737-223-9337
34	16	1.4	511	5	US-09-827-271-135
35	16	1.4	562	6	US-60-281-593-482
36	16	1.4	584	1	PCT-US01-01339-1755
37	16	1.4	584	1	PCT-US01-01339-1715
38	16	1.4	572	5	US-09-801-833-6254
39	16	1.4	647	1	PCT-US01-01339-8336
40	16	1.4	647	1	PCT-US01-01339-8338
41	16	1.4	675	4	US-08-956-171C-887
42	16	1.4	744	4	US-08-956-171C-966
43	16	1.4	1166	5	US-09-603-124B-433
44	16	1.4	1361	5	US-09-569-899-9
45	16	1.4	1599	5	US-09-739-449-5554
46	16	1.4	2114	1	PCT-US01-01302-178
47	16	1.4	2138	1	PCT-US01-01339-8314
48	16	1.4	2180	1	PCT-US01-01302-41
49	16	1.4	2225	1	PCT-US01-01301-359
50	16	1.4	2620	1	PCT-US01-01302-179
51	16	1.4	2663	5	US-09-654-935A-238
52	16	1.4	3167	1	PCT-US00-35017A-357
53	16	1.4	3710	5	US-09-543-679A-2477
54	16	1.4	4842	1	PCT-US01-01339-8504
55	16	1.4	4842	1	PCT-US01-01339-8505
56	16	1.4	4842	1	PCT-US01-01339-3032
57	16	1.4	4842	1	PCT-US01-01329-3033
58	16	1.4	8365	5	US-09-543-679A-3012
59	16	1.4	8414	1	PCT-US01-01312-1278
60	16	1.4	8987	6	US-60-248-505-628
61	16	1.4	11991	1	PCT-US01-01339-9893
62	16	1.4	13086	4	US-08-956-171C-16
63	16	1.4	14273	1	PCT-US01-01339-10125
64	16	1.4	15223	5	US-09-444-067-1
65	16	1.4	23694	6	US-60-248-505-601
66	16	1.4	24034	6	US-60-248-823-40
67	16	1.4	28080	6	US-60-248-505-70
68	16	1.4	32212	1	PCT-US01-01339-8770
69	16	1.4	32212	1	PCT-US01-01339-3299
70	16	1.4	36519	6	US-60-248-505-370
71	16	1.4	39789	6	US-60-248-505-157
72	16	1.4	49722	6	US-60-248-505-501
73	16	1.4	50193	6	US-60-248-505-351
74	16	1.4	50713	6	US-60-248-505-85
75	16	1.4	55193	6	US-60-248-505-511
76	16	1.4	70661	6	US-60-248-505-133
77	16	1.4	82769	6	US-60-248-823-20
78	16	1.4	106174	6	US-60-248-505-419
79	16	1.4	162025	5	US-09-834-700-13
80	16	1.4	162025	5	US-09-834-700-17
81	16	1.4	162025	5	US-09-834-700-14
82	16	1.4	162025	5	US-09-834-700-17
83	16	1.4	237961	6	US-09-834-700-18
84	16	1.4	249957	6	US-60-251-317-2
85	16	1.4	315341	5	US-09-335-032-12206
86	16	1.4	318095	5	US-09-739-449-215
87	16	1.4	334854	6	US-60-248-505-28
88	16	1.4	443925	6	US-60-248-505-299
89	16	1.4	443925	6	US-60-248-505-299
90	16	1.4	479874	6	US-60-248-505-72
91	16	1.4	582670	6	US-60-248-505-4
92	16	1.4	666448	5	US-09-335-032-12214
93	16	1.4	690793	6	US-60-248-505-227
94	16	1.4	784328	5	US-09-335-032-12217
95	16	1.4	948061	5	US-09-335-032-12219
96	16	1.4	1531974	5	US-09-335-032-12207
97	16	1.4	1531974	5	US-09-335-032-12207
98	15	1.3	125	5	US-09-540-212A-26471
99	15	1.3	128	5	US-09-540-212A-26471
100	15	1.3	151	5	US-09-487-566A-5297

Sequence 9895, Ap
Sequence 31535, A
Sequence 16219, A
Sequence 6168, Ap
Sequence 5608, Ap
Sequence 9337, Ap
Sequence 135, App
Sequence 482, App
Sequence 1755, Ap
Sequence 715, App
Sequence 6254, Ap
Sequence 8336, Ap
Sequence 838, Ap
Sequence 887, App
Sequence 966, App
Sequence 433, App
Sequence 9, Appl
Sequence 5554, Ap
Sequence 178, App
Sequence 8314, Ap
Sequence 41, Appl
Sequence 359, App
Sequence 179, App
Sequence 238, App
Sequence 357, App
Sequence 2477, Ap
Sequence 8504, Ap
Sequence 8505, Ap
Sequence 3032, Ap
Sequence 3033, Ap
Sequence 3033, Ap
Sequence 3012, Ap
Sequence 1278, Ap
Sequence 628, App
Sequence 9893, Ap
Sequence 16, Appl
Sequence 10125, A
Sequence 1, Appl
Sequence 601, App
Sequence 40, Appl
Sequence 70, Appl
Sequence 8770, Ap
Sequence 3299, Ap
Sequence 370, App
Sequence 157, App
Sequence 501, App
Sequence 351, App
Sequence 85, Appl
Sequence 511, App
Sequence 133, App
Sequence 20, Appl
Sequence 419, App
Sequence 13, Appl
Sequence 14, Appl
Sequence 17, Appl
Sequence 18, Appl
Sequence 2, Appl
Sequence 177, App
Sequence 12206, A
Sequence 215, App
Sequence 28, Appl
Sequence 299, App
Sequence 299, App
Sequence 72, Appl
Sequence 4, Appl
Sequence 12214, App
Sequence 227, App
Sequence 12217, A
Sequence 12219, A
Sequence 12207, A
Sequence 12207, A
Sequence 26471, A
Sequence 7725, Ap
Sequence 5297, Ap

C 539	1.2	14	4	US-08-956-171C-4060	Sequence 4060, Ap	612	14	1.2	424	1	PCT-US01-01339-6135	Sequence 6135, Ap
C 540	1.2	14	5	US-09-724-866A-7096	Sequence 7096, Ap	613	14	1.2	424	5	US-09-737-223-3819	Sequence 3819, Ap
C 541	1.2	14	5	US-09-689-912-433	Sequence 433, App	614	14	1.2	425	1	PCT-US01-01339-6136	Sequence 6136, Ap
C 542	1.2	14	1	PCT-US01-01351-32	Sequence 32, App	615	14	1.2	425	1	PCT-US01-01339-6137	Sequence 6137, Ap
C 543	1.2	14	5	US-09-724-866A-3948	Sequence 3948, Ap	616	14	1.2	425	5	US-09-801-833-2175	Sequence 2175, Ap
C 544	1.2	14	5	US-09-724-866A-6488	Sequence 6488, Ap	617	14	1.2	426	5	US-09-724-866A-2502	Sequence 2502, Ap
C 545	1.2	14	5	US-09-801-833-1007	Sequence 1007, Ap	618	14	1.2	426	5	US-09-724-866A-13996	Sequence 13996, A
C 546	1.2	14	5	US-09-724-866A-2225	Sequence 2225, Ap	619	14	1.2	427	5	US-09-724-866A-2789	Sequence 2789, Ap
C 547	1.2	14	5	US-09-724-866A-2233	Sequence 2233, A	620	14	1.2	428	4	US-08-276-163D-7604	Sequence 7604, Ap
C 548	1.2	14	5	US-08-956-171C-2980	Sequence 2980, Ap	621	14	1.2	428	4	US-09-724-866A-13996	Sequence 13996, A
C 549	1.2	14	5	US-09-724-866A-2007	Sequence 2007, Ap	622	14	1.2	429	5	US-09-737-223-7868	Sequence 7868, Ap
C 550	1.2	14	5	US-09-724-866A-13395	Sequence 13395, A	623	14	1.2	430	5	US-09-724-866A-2251	Sequence 2251, Ap
C 551	1.2	14	5	US-09-724-866A-24813	Sequence 24813, A	624	14	1.2	431	5	US-09-603-124B-275	Sequence 275, App
C 552	1.2	14	5	US-09-724-866A-19355	Sequence 19355, A	625	14	1.2	432	1	PCT-US00-35017A-90	Sequence 90, Appl
C 553	1.2	14	5	US-09-724-866A-22338	Sequence 22338, A	626	14	1.2	432	5	US-09-724-866A-20157	Sequence 20157, A
C 554	1.2	14	5	US-09-724-866A-23251	Sequence 23251, A	627	14	1.2	433	5	US-09-813-358-138	Sequence 138, App
C 555	1.2	14	5	US-09-724-866A-5634	Sequence 5634, Ap	628	14	1.2	434	5	US-09-724-866A-4449	Sequence 4449, Ap
C 556	1.2	14	5	US-09-724-866A-19071	Sequence 19071, A	629	14	1.2	434	5	US-09-737-223-14765	Sequence 14765, A
C 557	1.2	14	5	US-09-724-866A-22339	Sequence 22339, A	630	14	1.2	435	5	US-09-737-223-30909	Sequence 30909, A
C 558	1.2	14	5	US-09-737-223-6976	Sequence 6976, Ap	631	14	1.2	436	5	US-09-737-223-33897	Sequence 33897, A
C 559	1.2	14	5	US-09-724-866A-18570	Sequence 18570, A	632	14	1.2	436	5	US-09-737-223-17325	Sequence 17325, A
C 560	1.2	14	5	US-09-724-866A-19259	Sequence 19259, A	633	14	1.2	437	5	US-09-737-223-8864	Sequence 8864, Ap
C 561	1.2	14	5	US-09-724-866A-177	Sequence 177, App	634	14	1.2	438	5	US-09-724-866A-20998	Sequence 20998, A
C 562	1.2	14	5	US-09-724-866A-937	Sequence 937, App	635	14	1.2	439	5	US-09-724-866A-22654	Sequence 22654, A
C 563	1.2	14	5	US-09-724-866A-14034	Sequence 14034, A	636	14	1.2	440	5	US-09-737-223-13233	Sequence 13233, A
C 564	1.2	14	5	US-09-737-223-7675	Sequence 7675, Ap	637	14	1.2	440	5	US-09-737-223-9595	Sequence 9595, Ap
C 565	1.2	14	5	US-09-724-866A-2457	Sequence 2457, Ap	638	14	1.2	441	5	US-09-724-866A-5681	Sequence 5681, Ap
C 566	1.2	14	5	US-09-724-866A-13945	Sequence 13945, A	639	14	1.2	442	5	US-09-724-866A-1368	Sequence 1368, Ap
C 567	1.2	14	5	US-09-724-866A-16571	Sequence 16571, A	640	14	1.2	443	5	US-09-724-866A-1449	Sequence 1449, Ap
C 568	1.2	14	5	US-09-724-866A-15077	Sequence 15077, A	641	14	1.2	444	5	US-09-737-223-9032	Sequence 9032, Ap
C 569	1.2	14	5	US-09-833-263-459	Sequence 459, App	642	14	1.2	445	5	US-09-724-866A-5652	Sequence 5652, App
C 570	1.2	14	5	US-09-724-866A-2497	Sequence 2497, Ap	643	14	1.2	446	5	US-09-827-271-117	Sequence 117, App
C 571	1.2	14	5	US-09-737-223-3303	Sequence 3303, Ap	644	14	1.2	447	5	US-08-276-163D-5852	Sequence 5852, Ap
C 572	1.2	14	5	US-09-737-223-3343	Sequence 3343, Ap	645	14	1.2	448	5	US-09-737-223-33423	Sequence 33423, A
C 573	1.2	14	5	US-09-737-223-33192	Sequence 2215, Ap	646	14	1.2	449	5	US-09-737-223-11432	Sequence 11432, A
C 574	1.2	14	5	US-08-956-171C-2215	Sequence 2215, Ap	647	14	1.2	450	5	US-09-737-223-3797	Sequence 3797, Ap
C 575	1.2	14	5	US-09-737-223-6740	Sequence 6740, Ap	648	14	1.2	451	5	US-09-737-223-37242	Sequence 37242, Ap
C 576	1.2	14	5	US-09-801-833-5196	Sequence 5196, Ap	649	14	1.2	452	5	US-09-737-223-2024	Sequence 2024, Ap
C 577	1.2	14	5	US-09-724-866A-13561	Sequence 13561, Ap	650	14	1.2	453	5	US-09-724-866A-11571	Sequence 11571, A
C 578	1.2	14	5	US-09-737-223-17373	Sequence 17373, A	651	14	1.2	454	5	US-09-737-223-26388	Sequence 26388, A
C 579	1.2	14	5	US-09-724-866A-15004	Sequence 15004, A	652	14	1.2	455	5	US-09-737-223-24163	Sequence 24163, A
C 580	1.2	14	5	US-09-737-223-7447	Sequence 7447, Ap	653	14	1.2	456	5	US-09-724-866A-21030	Sequence 21030, A
C 581	1.2	14	5	US-09-737-223-11566	Sequence 11566, A	654	14	1.2	457	5	US-09-737-223-25560	Sequence 25560, A
C 582	1.2	14	5	US-09-724-866A-5491	Sequence 5491, Ap	655	14	1.2	458	5	US-09-813-358-207	Sequence 207, App
C 583	1.2	14	5	US-09-724-866A-22935	Sequence 22935, A	656	14	1.2	459	5	US-09-604-287A-1830	Sequence 1830, Ap
C 584	1.2	14	5	US-09-737-223-6446	Sequence 6446, Ap	657	14	1.2	460	5	US-09-640-211A-422	Sequence 422, App
C 585	1.2	14	5	US-09-737-223-4336	Sequence 4336, Ap	658	14	1.2	461	5	US-09-487-566A-4124	Sequence 4124, Ap
C 586	1.2	14	5	US-09-737-223-35844	Sequence 35844, A	659	14	1.2	462	5	US-09-689-912-772	Sequence 772, App
C 587	1.2	14	5	US-09-737-223-36156	Sequence 36156, A	660	14	1.2	463	5	US-09-724-866A-2943	Sequence 2943, Ap
C 588	1.2	14	5	US-09-737-223-17091	Sequence 17091, A	661	14	1.2	464	5	US-09-724-866A-15812	Sequence 15812, A
C 589	1.2	14	5	US-09-737-223-17583	Sequence 17583, A	662	14	1.2	465	5	US-09-724-866A-16057	Sequence 16057, A
C 590	1.2	14	5	US-09-737-223-17360	Sequence 1459, A	663	14	1.2	466	5	US-09-724-866A-20082	Sequence 20082, A
C 591	1.2	14	5	US-09-737-223-3143	Sequence 16583, A	664	14	1.2	467	5	US-09-737-223-9903	Sequence 9903, Ap
C 592	1.2	14	5	US-09-737-223-17692	Sequence 17692, A	665	14	1.2	468	5	US-09-724-866A-13100	Sequence 13100, A
C 593	1.2	14	5	US-09-724-866A-641	Sequence 641, App	666	14	1.2	469	5	US-09-724-866A-17908	Sequence 17908, A
C 594	1.2	14	5	US-09-737-223-35885	Sequence 35885, A	667	14	1.2	470	5	US-09-737-223-29319	Sequence 29319, A
C 595	1.2	14	5	US-09-737-223-35642	Sequence 1965, Ap	668	14	1.2	471	5	US-09-737-223-5472	Sequence 5472, Ap
C 596	1.2	14	5	US-09-533-077-279	Sequence 35642, A	669	14	1.2	472	5	US-09-724-866A-23670	Sequence 23670, A
C 597	1.2	14	5	US-09-724-866A-15428	Sequence 15428, A	670	14	1.2	473	5	US-09-737-223-12109	Sequence 12109, A
C 598	1.2	14	5	US-09-737-223-33658	Sequence 279, App	671	14	1.2	474	5	US-09-737-223-13145	Sequence 13145, A
C 599	1.2	14	5	US-09-737-223-3143	Sequence 33588, A	672	14	1.2	475	5	US-09-737-223-14029	Sequence 14029, A
C 600	1.2	14	5	US-09-737-223-17692	Sequence 3143, Ap	673	14	1.2	476	5	US-09-737-223-14018	Sequence 14018, A
C 601	1.2	14	5	US-09-724-866A-641	Sequence 17692, A	674	14	1.2	477	5	US-09-737-223-30857	Sequence 30857, A
C 602	1.2	14	5	US-09-737-223-35885	Sequence 641, App	675	14	1.2	478	5	US-09-724-866A-23090	Sequence 23090, A
C 603	1.2	14	5	US-09-801-833-6862	Sequence 35885, A	676	14	1.2	479	5	US-09-737-223-717	Sequence 717, App
C 604	1.2	14	5	US-09-737-223-16828	Sequence 6862, Ap	677	14	1.2	480	5	US-09-724-866A-13123	Sequence 13123, A
C 605	1.2	14	5	US-09-801-833-1230	Sequence 1230, Ap	678	14	1.2	481	5		
C 606	1.2	14	5	US-09-737-223-5869	Sequence 5869, Ap	679	14	1.2	482	5		
C 607	1.2	14	5	US-09-640-211A-1625	Sequence 1625, Ap	680	14	1.2	483	5		
C 608	1.2	14	5	PCT-US01-01339-1765	Sequence 1765, App	681	14	1.2	484	5		
C 609	1.2	14	5	US-09-724-866A-18078	Sequence 725, App	682	14	1.2	485	5		
C 610	1.2	14	5	US-09-724-866A-20138	Sequence 18078, A	683	14	1.2	486	5		
C 611	1.2	14	5	US-09-737-223-5775	Sequence 20138, A	684	14	1.2	487	5		
			4	US-08-276-163D-2966	Sequence 2966, Ap							

Tue May 15 07:26:57 2001

us-09-151-409-15.oligo.rnpn

Sequence 2784, Ap
Sequence 2790, Ap
Sequence 7, Appl
Sequence 18, Appl
Sequence 9644, Ap
Sequence 553, App
Sequence 312, App
Sequence 392, App
Sequence 587, App
Sequence 393, App
Sequence 358, App
Sequence 8780, Ap
Sequence 3309, Ap
Sequence 409, App
Sequence 8781, Ap
Sequence 3310, Ap
Sequence 6624, Ap
Sequence 822, App
Sequence 2377, Ap
Sequence 551, App
Sequence 531, App
Sequence 9640, Ap
Sequence 9794, Ap
Sequence 332, App

14 1.2 4249 1 PCT-US01-01329-2784
14 1.2 4254 1 PCT-US01-04098A-2790
14 1.2 4286 5 US-09-817-856-7
14 1.2 4305 1 US-09-846-351-18
14 1.2 4370 1 PCT-US01-01339-9644
14 1.2 4399 1 PCT-US01-04098A-553
14 1.2 4446 6 US-60-281-593-312
14 1.2 4694 1 PCT-US01-04098A-392
14 1.2 4709 4 US-08-956-171C-587
14 1.2 4739 1 PCT-US01-04098A-393
14 1.2 4747 1 PCT-US00-35017A-358
14 1.2 4800 1 PCT-US01-01339-8780
14 1.2 4800 1 PCT-US01-01329-3309
14 1.2 4844 1 PCT-US01-04098A-409
14 1.2 4845 1 PCT-US01-01339-8781
14 1.2 4859 1 PCT-US01-01339-6624
14 1.2 4862 1 PCT-US01-04098A-822
14 1.2 4879 1 PCT-US01-04098A-2377
14 1.2 4923 4 US-08-956-171C-551
14 1.2 4945 6 US-60-248-505-531
14 1.2 4964 5 US-09-801-833-7640
14 1.2 5086 1 PCT-US01-01339-9794
14 1.2 5136 4 US-08-956-171C-332
c1000

ALIGNMENTS

RESULT 1
US-09-540-212A-66707
; Sequence 66707, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Sellhame, Jeffrey J.
; APPLICANT: Deleageane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 66707
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No: hu00684568
US-09-540-212A-66707

Query Match 1.6%; Score 18; DB 5; Length 275;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 taactgaagagttgagta 169
|||||
DB 169 taactgaagagttgagta 186

RESULT 2
US-08-956-171C-537/c
; Sequence 537, Application US/08956171C
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash

Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171C
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Hoover, Kenley K.
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 610-5790
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 537:
SEQUENCE CHARACTERISTICS:
LENGTH: 784 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 537:
US-08-956-171C-537

Query Match 1.6%; Score 18; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 aactttttaacaagtag 439
|||||
DB 673 AACTTCTTAAACAAGTAG 656

RESULT 3
PCT-US01-04098A-471
; Sequence 471, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20

```
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 471
; LENGTH: 4813
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (155)..(4813)
PCT-US01-04098A-471

Query Match      1.6%; Score 18; DB 1; Length 4813;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 cagaatgagaagttatct 591
    |||||
Db 245 cagaatgagaagttatct 262

RESULT 4
PCT-US01-04098A-2439/c
; Sequence 2439, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 2439
; LENGTH: 5623
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01-04098A-2439

Query Match      1.6%; Score 18; DB 1; Length 5623;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 cagaatgagaagttatct 591
    |||||
Db 5419 CAGAAAGAGAGTATCT 5402

RESULT 5
US-60-248-505-134
; Sequence 134, Application US/60248505
```

```
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: G1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 283061
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(283061)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-505-134

Query Match      1.6%; Score 18; DB 6; Length 283061;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1113 aaaagctagtaaaattca 1130
    |||||
Db 280582 aaaagctagtaaaattca 280599

RESULT 6
US-09-540-212A-13409/c
; Sequence 13409, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 13409
; LENGTH: 132
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00300062
; NAME/KEY: unsure
; LOCATION: 118
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-212A-13409

Query Match      1.5%; Score 17; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 519 taagctaaaaaattta 535
    |||||
Db 81 TAAGCTAAAAAATTa 65

RESULT 7
US-09-801-833-7074/c
; Sequence 7074, Application US/09801833
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
```

```

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; TITLE OF INVENTION: HUMAN BRAIN LIBRARY
; FILE REFERENCE: 1600.1037-005
; CURRENT APPLICATION NUMBER: US/09/801,833
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/371,168
; PRIOR FILING DATE: 1999-08-10
; PRIOR APPLICATION NUMBER: 60/095,907
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/103,145
; PRIOR FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8285
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7074
; LENGTH: 1682
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1682)
; OTHER INFORMATION: n = A,T,C or G
US-09-801-833-7074

Query Match
Best Local Similarity 100.0%; Score 17; DB 5; Length 1682;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 765 ggtcacacaattatata 781
DB 1315 GGTACACAATTATATA 1299

RESULT 8
US-09-543-679A-2639/c
; Sequence 2639, Application US/09543679A
; GENERAL INFORMATION:
; APPLICANT: NYCE, Jonathan W.
; TITLE OF INVENTION: LOW ADENOSINE ANTI-SENSE OLIGONUCLEOTIDE,
; COMPOSITIONS, KIT & METHOD FOR TREATMENT
; OF AIRWAY DISORDERS ASSOCIATED WITH
; BRONCHOCONSTRICTION, LUNG INFLAMMATION,
; NUMBER OF SEQUENCES: 3111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: EPIGENESIS PHARMACEUTICALS, INC.
; STREET: 7 Clarke Drive
; CITY: Cranbury
; STATE: NJ
; COUNTRY: USA
; ZIP: 08512
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,679A
; FILING DATE: 13-Apr-2000
; CLASSIFICATION: UNKNOWN
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/127,958
; FILING DATE: 1998-08-03
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: RPI-0067191b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-409-3035
; TELEFAX: 413-254-9245
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2639:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3929 base pairs

US-09-151-409-15.oligo.rnpn

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2639
US-09-543-679A-2639

Query Match
Best Local Similarity 100.0%; Score 17; DB 5; Length 3929;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 aaactaagaatgaag 293
DB 1513 AAAACTAAGAATGAAG 1497

RESULT 9
US-08-956-171C-37
; Sequence 37, Application US/08956171C
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171C
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Kenley K.
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9834 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-08-956-171C-37

Query Match
Best Local Similarity 100.0%; Score 17; DB 4; Length 9834;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 taataataaagcgttaa 133
DB 3593 TAATAATAAGCGTTAA 3609
```

```
RESULT 10
US-09-543-679A-2640/c
; Sequence 2640, Application US/09543679A
; GENERAL INFORMATION:
; APPLICANT: NYCE, Jonathan W.
; TITLE OF INVENTION: LOW ADENOSINE ANTI-SENSE OLIGONUCLEOTIDE,
; COMPOSITIONS, KIT & METHOD FOR TREATMENT
; OF AIRWAY DISORDERS ASSOCIATED WITH
; BRONCHOCONSTRICTION, LUNG INFLAMMATION,
; NUMBER OF SEQUENCES: 3111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: EPIGENESIS PHARMACEUTICALS, INC.
; STREET: 7 Clarke Drive
; CITY: Cranbury
; STATE: NJ
; COUNTRY: USA
; ZIP: 08512
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,679A
; FILING DATE: 13-Apr-2000
; CLASSIFICATION: UNKNOWN
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/127,958
; FILING DATE: 1998-08-03
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: EPI-0067191b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-409-3035
; TELEFAX: 413-254-9245
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2640:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2640
US-09-543-679A-2640

Query Match 1.5%; Score 17; DB 5; Length 12447;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 aaactaagaatgaag 293
|||||
Db 10031 AAACTAAGATGAAGG 10015

RESULT 11
US-60-248-505-614/c
; Sequence 614, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 614
; LENGTH: 126883
```

```
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(126883)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-505-614

Query Match 1.5%; Score 17; DB 6; Length 126883;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 ttaagttttaataataa 125
|||||
Db 26637 TTAAGTTTAAATAATAA 26621

RESULT 12
US-09-335-032-12216
; Sequence 12216, Application US/09335032
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 01107.78572
; CURRENT APPLICATION NUMBER: US/09/335,032
; CURRENT FILING DATE: 1999-06-16
; PRIOR APPLICATION NUMBER: US 60/035,917
; PRIOR FILING DATE: 1997-01-23
; PRIOR APPLICATION NUMBER: US 09/012,031
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 12219
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12216
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-335-032-12216

Query Match 1.5%; Score 17; DB 5; Length 924430;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 aattacaacaagaat 579
|||||
Db 11585 aattacaacaagaat 115601

RESULT 13
US-09-540-212A-37688/c
; Sequence 37688, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Seilhame, Jeffrey J.
; APPLICANT: Deleage, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullish, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 37688
; LENGTH: 95
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00858432
US-09-540-212A-37688

Query Match 1.4%; Score 16; DB 5; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 519 taagctaaaaaaatt 534
|||||
Db 68 TAAGCTAAAAAAATT 53

RESULT 14
US-09-540-212A-21769/c
; Sequence 21769, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Sellhammer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 21769
; LENGTH: 120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00384850
US-09-540-212A-21769

Query Match 1.4%; Score 16; DB 5; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 taaaaaactaagaatg 289
|||||
Db 45 TTAATAAACTAAGAATG 30

RESULT 15
US-09-815-343-1168/c
; Sequence 1168, Application US/09815343
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1168
; LENGTH: 165
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-1168

Query Match 1.4%; Score 16; DB 5; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 cggatgatggtatcct 615
|||||
Db 66 CGGATGATGGTAATCCT 51

Search completed: May 13, 2001, 06:26:17
Job time: 3396 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:07 ; Search time 44.05 Seconds
(without alignments)
1398.576 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 1954
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLERASKIQELEAR 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1009251 seqs, 150854530 residues

Total number of hits satisfying chosen parameters: 1009251

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1954	100.0	383	15	US-09-151-409-16
2	312	16.0	443	7	US-08-325-278-6
3	256	13.1	440	17	US-09-302-756-35
4	220.5	11.3	327	11	US-08-761-184-1230
5	220.5	11.3	327	11	US-08-761-184-1230
6	220.5	11.3	327	12	US-08-821-931-1230
7	220.5	11.3	327	12	US-08-821-931-1230
8	220.5	11.3	327	13	US-08-993-002A-8616
9	220.5	11.3	327	13	US-08-993-002A-8617
10	219	11.2	810	16	US-09-248-796-20281, A

11	218.5	11.2	1078	16	US-09-248-796-20284
12	215.5	11.0	630	16	US-09-248-796-20275
13	212	10.8	1193	23	US-60-167-217-2860
14	212	10.8	1193	23	US-60-173-464-2336
15	212	10.8	1457	23	US-60-191-637-2817
16	212	10.8	1457	23	US-60-191-681-2256
17	207.5	10.6	980	1	PCT-US99-26796-152
18	206.5	10.6	2349	1	PCT-US99-17738-68
19	203	10.4	42	1	PCT-US94-12268-31
20	203	10.4	42	5	US-08-143-412A-31
21	203	10.4	42	8	US-08-479-690-31
22	201	10.3	905	16	US-09-248-796-16333
23	200	10.2	1939	17	US-09-310-187-1
24	200	10.2	1939	17	US-09-310-187A-1
25	198.5	10.2	2954	15	US-09-150-867-1
26	198.5	10.2	2954	19	US-09-536-933-1
27	196	10.0	1948	23	US-60-259-128-4790
28	196	10.0	1979	1	PCT-US99-26796-28
29	195.5	10.0	1788	18	US-09-488-725A-5398
30	195	10.0	1381	23	US-60-191-637-552
31	195	10.0	1381	23	US-60-191-681-435
32	194.5	10.0	2688	18	US-09-488-725A-5814
33	193.5	9.9	561	13	US-08-967-909-2
34	193	9.9	722	16	US-09-248-796-20613
35	191.5	9.8	2663	18	US-09-488-725A-2242
36	190.5	9.7	753	18	US-09-458-113-2
37	190.5	9.7	1690	23	US-60-167-217-10243
38	190.5	9.7	1690	23	US-60-167-217-10339
39	190.5	9.7	1690	23	US-60-173-464-8250
40	190.5	9.7	1690	23	US-60-173-464-8335
41	190.5	9.7	1690	23	US-60-191-637-10256
42	190.5	9.7	1690	23	US-60-191-637-10343
43	190.5	9.7	1690	23	US-60-191-681-7997
44	190.5	9.7	1690	23	US-60-191-681-8069
45	190.5	9.7	1780	18	US-09-488-725A-1826

ALIGNMENTS

RESULT 1

US-09-151-409-16
; Sequence 16, Application US/09151409A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: GROUP A STREPTOCOCCAL VACCINES
; FILE REFERENCE: 481112.410
; CURRENT APPLICATION NUMBER: US/09/151,409A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,635
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Hexavalent M
; OTHER INFORMATION: fusion gene sequence constructed from
; OTHER INFORMATION: streptococcal type 24, 5, 6, 19, 1 and 3 M protein
; OTHER INFORMATION: DNAs
US-09-151-409-16

Query Match 100.0%; Score 1954; DB 15; Length 383;
Best Local Similarity 100.0%; Pred. No. 6e-127;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACWATRSQDTLEKVOERADKFEIENNTLKNSDLSPNNKALKDHNDLETELSNAKE 60
|||||
Db 1 ACWATRSQDTLEKVOERADKFEIENNTLKNSDLSPNNKALKDHNDLETELSNAKE 60

Sequence 20284, A
Sequence 20275, A
Sequence 2860, Ap
Sequence 2336, Ap
Sequence 2817, Ap
Sequence 2256, Ap
Sequence 152, App
Sequence 68, Appl
Sequence 31, Appl
Sequence 31, Appl
Sequence 16333, A
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 28, Appl
Sequence 5398, Ap
Sequence 552, App
Sequence 435, App
Sequence 5814, Ap
Sequence 2, Appli
Sequence 20613, A
Sequence 2242, Ap
Sequence 2, Appli
Sequence 10243, A
Sequence 10339, A
Sequence 8260, Ap
Sequence 8335, Ap
Sequence 10256, A
Sequence 10343, A
Sequence 7997, Ap
Sequence 8069, Ap
Sequence 1826, Ap

QY 253 LDQVTLQYTHKNSYQY-----NAQAGRLDLROKAEYLKGLNDWAER----- 295
DB 61 EDQKQLETKELQOQYDLAKESTSWDRQLEKELEKEKEALELAIDQASRDYHRATL 120
QY 296 -----LQELNIDVATR--SQTDLEKQVQERADK--FEIENNTL---KLKNSDLSPNN 341
DB 121 EKELEEKKALELAIDQASQDYNRANVLEKELETITREQEINRNLLGNNAKLELDQLSSEK 180
QY 342 KALKDHDELTEE--LSNAKEKLRKNDKSLSEKASK 375
DB 181 EQLTIERAKLEEEKQISDASRSQSLRRDLDSAREAKK 216
RESULT 3
US-09-302-756-35
; Sequence 35, Application US/09302756
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; TITLE OF INVENTION: THE SURFACE OF GRAM POSITIVE BACTERIA
; FILE REFERENCE: 016921-076
; CURRENT APPLICATION NUMBER: US/09/302,756
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: US 07/522,440
; EARLIER FILING DATE: 1990-05-11
; EARLIER APPLICATION NUMBER: US 07/742,199
; EARLIER FILING DATE: 1991-08-05
; EARLIER APPLICATION NUMBER: US 07/814,823
; EARLIER FILING DATE: 1991-12-23
; EARLIER APPLICATION NUMBER: US 07/951,082
; EARLIER FILING DATE: 1992-03-13
; EARLIER APPLICATION NUMBER: PCT/US93/02355
; EARLIER FILING DATE: 1993-03-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-09-302-756-35
Query Match 13.1%; Score 256; DB 17; Length 440;
Best Local Similarity 30.8%; Pred. No. 1.1e-09;
Matches 100; Conservative 34; Mismatches 111; Indels 80; Gaps 12;
QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANDKLPWRVRYTRHTPEDKLLIIDDLDK 185
DB 1 RVFPRGTVENPDKARELLNKYDVNSMLQANDKLT-----TENNLTQNKNLTE 52
QY 186 EHELOQNEKLSLQNGDGNPREVIEDLAANNPAI--QNIRLHENKDLKARLENAMVAG 243
DB 53 NKULTDQNKNLTTEN--KNLTQNKNLTTENKELKAEGRNLTNTENKGLTKLSEAEPEAA 110
QY 244 ---RDFKRA--GTLDDQVTLQYTHKNSYQYNAQAGRL-----DLROKAEYLKGLN 290
DB 111 NKERENKEAIGTLTKTLDVTVKDKIAKEQESKETIGTLKTLDETIVKDKIAKEQESKETI 170
QY 291 DWAEERLQELNIDVATRSTQTD-----TLKV-----QERADK 322
DB 171 GTLTKTLDVTVKDKIAKEQESKETIGTLKTLDETIVKDKIAKEQSKQDICALKQELAKK 230
QY 323 FE---TENNTLKLKNSDLSPNNKALKDHDE---LTEELSNKAK-----LRKN- 365
DB 231 DEGNKVSASRKLRLRRDLDSAREAKKQVEKDLANLTAEKDKYKKEKQISDASRQGLRRDL 290
QY 366 -----DKSLSEKASKIOELE 380
DB 291 DASREAKKQVEKALEEANSKLALE 315

QY 61 KLRSVAVTRGTINDPQAKALDKYLELHDLTKTNEGLKTENEGTKTENEGL 120
DB 61 KLRSVAVTRGTINDPQAKALDKYLELHDLTKTNEGLKTENEGTKTENEGL 120
QY 121 KTEVDVFPRTGTVENPDKARELLNKYDVNSMLQANDKLPWRVRYTRHTPEDKLLKIID 180
DB 121 KTEVDVFPRTGTVENPDKARELLNKYDVNSMLQANDKLPWRVRYTRHTPEDKLLKIID 180
QY 181 DLDAKEHELOQNEKLSLQNGDGNPREVIEDLAANNPAIONIRLHENKDLKARLENAM 240
DB 181 DLDAKEHELOQNEKLSLQNGDGNPREVIEDLAANNPAIONIRLHENKDLKARLENAM 240
QY 241 VAGROFKRAGTLDDQVTLQYTHKNSYQYNAQAGRLDLROKAEYLKGLNDWAERLLQEL 300
DB 241 VAGROFKRAGTLDDQVTLQYTHKNSYQYNAQAGRLDLROKAEYLKGLNDWAERLLQEL 300
QY 301 NIDVATRSTQTDLEKQVQERADKFEIENNTLKLKNSDLSPNNKALKDHDELTEELSNK 360
DB 301 NIDVATRSTQTDLEKQVQERADKFEIENNTLKLKNSDLSPNNKALKDHDELTEELSNK 360
QY 361 KLKNDKSLSEKASKIOELEAKK 383
DB 361 KLKNDKSLSEKASKIOELEAKK 383
RESULT 2
US-08-325-278-6
; Sequence 6, Application US/08325278
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/325,278
; FILING DATE: 26-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McWaters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 450023.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-278-6
Query Match 16.0%; Score 312; DB 7; Length 443;
Best Local Similarity 43.1%; Pred. No. 1.5e-13;
Matches 93; Conservative 18; Mismatches 65; Indels 40; Gaps 7;
QY 200 NCDGNPREVIEDLAANNPAIONIRLHENKDLKARLENAMVAGRDFKRA-----GTL 252
DB 1 NCDGNPREVIEDLAANNPAIONIRLHENKDLKARLENAMVAGRDFKRAEELKAKQAL 60

```

RESULT      4
US-08-761-184-1230
; Sequence 1230, Application US/08761184
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
; TITLE OF INVENTION: THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 1810
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/761,184
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1230:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...327
US-08-761-184-1230

Query Match          11.3%; Score 220.5; DB 11; Length 327;
Best Local Similarity 25.7%; Pred. No. 1.9e-07;
Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;

QY 14 EKVOERADFFETENTLKLNKSDLSFNNAKDHDEL---TEELSSNAKEIKLRGSAVTRG 70
|::: | : | | | : | : | | | | | : | | |
Db 16 EELEARISELDENTELLRRERYLAETSFKDANDQLRGKNDFITDKDL----- 67
|::: | : | | | : | : | | | | | : | | |
QY 71 TINDPQRAEKALDYKLENHDLTKNNEGLKTENEGLKTENEGLKTENEGLKTENVDRVTPR 130

```

	INFORMATION FOR SEQ ID NO:	1445;		
:	SEQUENCE CHARACTERISTICS:			
:	LENGTH: 327 amino acids			
:	TYPE: amino acid			
:	TOPOLOGY: linear			
:	MOLECULE TYPE: protein			
:	HYPOTHETICAL: YES			
:	ORIGINAL SOURCE:			
:	ORGANISM: Helicobacter pylori			
:	FEATURE:			
:	NAME/KEY: misc_feature			
:	LOCATION: 1...327			
:	US-08-761-184-1445			
	Query Match	11.3%; Score 220.5; DB 11; Length 327;		
	Best Local Similarity	25.7%, Pred. No. 1.9e-07;		
	Matches	97; Conservative	64; Mismatches	133; Indels
	Gaps	83;		
	Length	327;		
	DB	15;		
Qy	14 EKVOERADKFEIENNTLKLNDSLSFNKKALDKHDEL---TEELSNAKEKLRSVAVTRG	70		
Dd	16 EELEARISELEDENTELLRREREYLAATSELDANDQLRQNDKLFITKDRL-----	67		
Qy	71 TINDPQAKALKDYLENHDLTKNEGTKTENGLKTENEGLKTENEGLKTENVDRVPFR	130		
Dd	68 -----TKENTELF-AENESLVSIGTSLHSNDOQMNNKLTKEKAELKTE-----	112		
Qy	131 GTVENPDKARELLNKYDVNSMLQAANDKLPWRVRYTHRTPTDCLKKIIDDLDAKEHELQ	190		
Dd	113 -----KDILAK---ENTRLLAARDRLTEEKRELT--TEKERLKRENTTELTHKITELT	159		
Qy	191 QONEKLSLONGDNPREVIDLANNPATONIRLRHENKDKLARLENAMAVEAGDFKRAG	250		
Dd	160 KENKALTEN-----DKLNHQVTALTNERDSLQE--RARLODA-----HG	198		
Qy	251 TLDDOVTOLTYTKHNSNYQQNAOGRDLDRQAEYLKGLNDWAERLLQELNLIDVATRST	310		
Dd	199 FLEKRCNL-EKEN---QRILTDLKLESQAKS--LNTNNQLROALENSVOLAAQAKEK	252		
Qy	311 DTLEKVQERADKFIEINTTLKLN-----SDLSNNKALDKDHDELTEELSNAKEK---	361		
Dd	253 IAIE-----KSELEREIARKLSLEGMEAKSDLDLHNRLASANEIDLKRONRKLEEINI	305		
Qy	362 -LRNPKSLSKASKIQ 377			
Dd	306 ALKERVGDGNLSQLQ 322			
RESULT	6			
US-08-821-931-1230				
: Sequence 1230.	Application US/08821931			
: GENERAL INFORMATION:				
: APPLICANT:	DOUGLAS SMITH ET AL			
: TITLE OF INVENTION:	NUCLEIC ACID AND AMINO ACID SEQUENCES			
: RELATING TO:	RELATIONSHIP BETWEEN PLYORI AND			
: TITLE OF INVENTION:	THERAPEUTIC USES THEREOF			
: NUMBER OF SEQUENCES:	1810			
: CORRESPONDENCE ADDRESS:				
: ADDRESSEE:	LAHIVE & COCKFIELD			
: STREET:	60 State Street, Suite 510			
: CITY:	Boston			
: STATE:	Massachusetts			
: COUNTRY:	USA			
: ZIP:	02109-1875			
: COMPUTER READABLE FORM:				
: MEDIUM TYPE:	CD-ROM ISO9660			
: COMPUTER:				
: OPERATING SYSTEM:				
: SOFTWARE:				
: CURRENT APPLICATION DATA:				
: APPLICATION NUMBER:	US/08/821.931			
: FILING DATE:				
: PRIOR APPLICATION DATA:				

```

RESULT 7
US-08-821-931-1445
; Sequence 1445, Application US/08821931
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
; TITLE OF INVENTION: THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 1810
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/821,931
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/761,184
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP9CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1445:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...327
US-08-821-931-1445

```

Query Match 11.3%; Score 220.5; DB 12; Length 327;
 Best Local Similarity 25.7%; Pred. No. 1.9e-07;
 Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;

```

QY 14 EKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDEL---TELSNAKEKLRGSAVTRG 70
Db 16 EELEARISEDENTELLRREREYLAETSELKDANDQLRKNDKLFITDKL----- 67
QY 71 TINDPQRAKEALDKYELENDHDKTKNEGLKTENEGTKTENEGTKTEVDRVFPFR 130
Db 68 -----TKENTELF-AENESLSVKISGLEHSDQWNNKLTKEAEKTE----- 112
QY 131 GTVENPKARELLKNDYVENSMLQANNDKLPWRVRYTRTPEDKCLKKIIDDLDAAKEHLQ 190
Db 113 -----KDILAK---ENTRLAARDRLTEKRELT-TEKERLKRNTLTHITELT 159
QY 191 QONEKLISLONGDGNPREVIEDLAANPAONIRLAKHNKDLKARLENAMAEVAGRDFKRAK 250
Db 160 KENKALTEN-----DKLNHQVTALTNERDSLEQE--RARIQDA-----HG 198
QY 251 TLDDQVTQLYTKHNSYQYNAQAGRLDLROKAEYLKGLNDWAERLLQELNIDVATRSQT 310
Db 199 FLEKRCNLI-EKEN---QRLTDKLLQLESQAQS--LENTNQLQALENSINVQLAAQAEK 252
QY 311 DTLEKVOERADKFEIENNTLKLKN-----SDLSFNKALKDHNDELTELSNAKEK-- 361
Db 253 IAE-----KSELERETARUKSLEGMKASDLDLHNRRLASANEDLKQRNKRLEENI 305
QY 362 -LRKNDKLSLSEKASKIQ 377
Db 306 ALKERVDTGLNEQLSKIQ 322

RESULT 8
US-08-993-002A-8616
; Sequence 8616, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8616:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:

```


Qy	1	ACMVATRSQTDTL-----EKVQERADKFEIENNTLKNSDLSFNKA-----LKDHNDLT 52
Db	436	SAMKASKKEFETASQYQRELOEALFKGWTSESTLKQLEKLDSTEQAKKKLELDGINMT 495
Qy	53	EELSNAKEXLGRSAVTRGTINDPORAKEALD-KYELEN-----HDLTKNEGLKTE 102
Db	496	RDLFHLK-LKLSAETQ----IKOREREFKNLTYELENTKDYELQINLNSNNEFKOK 550
Qy	103	NEGLKTENEGLKTENGLKTEVDVFPFGTVENPKARELLNKYDVE-NSMLQANNDKLP 161
Db	551	INELSKKIESLTEDNKSNAQLEELK-RTENENNEHMDKLRASVAYNDLKAKASEEE 609
Qy	162	WRVYTRHTPEBKLLIIDDLAKAHEHQOQNEKS-----LQNGDGNPREVIEDL---- 212
Db	610	ETVK-----AKEELETLTISKIDNELKELKEQOQSKNLEBGLQNLITDSTNEKFKLEDEL 664
Qy	213	-----AANPAQNIRLHENKDLKARLNAMEVAGRDPKFRAGTLTLLDQVTOPLY 261

[illegible]

RESULT 13
US-60-167-217-2860
Sequence 2860, Application US/60167217

```

; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS. NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2860
; LENGTH: 1193
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1193)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-167-217-2860

Query Match 10.8%; Score 212; DB 23; Length 1193;
Best Local Similarity 26.2%; Pred. No. 5e-06;
Matches 110; Conservative 66; Mismatches 176; Indels 68; Gaps 16;

QY 7 RSQDTLEKVGQERADKFEIENNTLK--LKNSDLFSFNKALKDHNDLTELSELSNAEKL-- 62
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 269 RSSAD-WESTKQRIARLENERLKHDLERSQTTFGRTM-----TTSQELDRAQERADK 322
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 63 -----RGSAVTGTINDPORAK-----EALQYELHNHDLTKNEGLKTENEGKLTE 109
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 323 ASAE LRRTQAE LKVTQSDAERAEAAALQEKLEKSGEYVRLKAKLENAQGEQSLRQE 382
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 110 NEGLKTENEGKLTEDVRVP--RGTVENPDKARELLNKYDVENSMLQANDKLPWVRVT 167
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 383 LEKAQGSVRIHADRDRAFESEVEKIEEMERTQATLGKSLQHEKLSLDKKAQNEVDHL 442
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 168 RHPEDKL-----KKIIDLDAKEHEL-----QQQNEKLSLQNGDG 203
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 443 ----QDKLDKACTENRRLVEKELTYDNLQSQLDKALGQAARMQKERETLSLDT--D 496
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 204 NPREVIEDLAANPAQNTNR--LRHNKDLKARLENAMEV--AGROFKRAGTLLIDQVTQ 258
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 497 RIREKLEKTQVGLGRIGKQERDQFSDELTGKRSQAOTLLMKAAARDREAMQTDLEVIKE 556
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 259 LYTKHNSYQYNAQAGRLDLROKAYLGLNDWAERLLQELINIDVATRSQTDTLKRVQE 318
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 557 RYRK--SHAIQOKLQMERDDAVTEVELKEKLDKALYASQKL-IDEKDTSNKE-FEKMLE 612
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 319 RADKFEIENNTLKLNKSDLSFNKALKDHNDLTELSELSNAEKLKRNKDKSLSEKASKIQE 378
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 613 KYDRAQNEIYRLQSKCDTAEDRARLEVEAERSGLAASKAREDLRK----LQDESTRLQE 668
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 14
US-60-173-464-2336
; Sequence 2336, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL0000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2336
; LENGTH: 1193
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT

```

```

; LOCATION: (1)...((1193)
; OTHER INFORMATION: xaa = Any Amino Acid
US-60-173-464-2336

Query Match 10.8%; Score 212; DB 23; Length 1193;
Best Local Similarity 26.2%; Pred. No. 5e-06;
Matches 110; Conservative 66; Mismatches 176; Indels 68; Gaps 16;

QY 7 RSOTDTLEKVVQERADRFETENNLT--LKNSDLSFNNKALKDHNDLTELTEELSNAKEKL-- 62
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 269 RSSAD-WESTKQRIARLELENERLKHDLERSQTTFGRTM-----TTSOELDRAQERADK 322

QY 63 -----RGSVTTGTINDPORAK-----EALDKYELFNHDLKTKYNEGLKTENGLKTE 109
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 323 ASABLRTQAEIQLVTSQDAERAREEAAAQLEKLESGOGEVIRLKALENAQGEQESLROE 382

QY 110 NEGLRTNEGLKTEVDRFP--RGTVENPDKARELLNKYDVNSMLQANDKLPNRRVRYT 167
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 383 LEKAQSGVSRIHADRDRAFSEFKIEEMERTQATLGKSOLQHEKLNQSLDKRAQNEVDHL 442

QY 168 RHTPEDKL-----KKIIDDLDKAKEHEL-----QQQNEKLSLQNGDG 203
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 443 ----QDKLDRACTENRRLLVLEKEKLTVDYDNLQSLDKALGQAARMQKERETLSLDT--D 496

QY 204 NPREVIEDLAANPAQINR--LRHENKDLKARLENAMEV---AGRDFKRAGLTLLDQVQT 258
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 497 RIREKLEKTQVLGRIQERQDFSELETLKERSESAQTLLMKAARDREAMQTDLEVLKE 556

QY 259 LYTKHNSVQYNAQAGRLDLRQKAEVLKGLNDWAERLLQELNIDVATRSQTDLTLEKVQE 318
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 557 RYEK--SHAIQOKLQMERDDAVTEVILKEKLDKALYASQKL-IDEKDTSNKE-FEKMLE 612

QY 319 RADRFENNTLKLKNSDLSFNNKALKDHNDLTELTEELSNAKEKLKNDKSLSKASKIQE 378
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 613 KYDRAQNEIYRLQSRCDTAEDRARLEAVEERSGLAAKAKAREDLURK-----LQDESTR 668

RESULT 15
US-60-191-637-2817
; sequence 2817, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191.637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2817
; LENGTH: 1457
; TYPE: PRF
; ORGANISM: DROSOPHILA
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1457)
; OTHER INFORMATION: xaa = Any Amino Acid
US-60-191-637-2817

```

```

Query Watch      10.8%   Score 212; DB 23; Length 1457;
Best Local Similarity 26.2%; Pred. No. 6.7e-06;
Matches 110; Conservative 66; Mismatches 176; Indels 68; Gaps 16;

QY       7 RSOTDTLEKVVQERADKFIEIENNTK--LKNSDLSFNNKALKDHNDLTETELSSNAKEKL-- 62
          ||| :|| :|| ||| - - - - - :||| :|||
Db       269 RSSAD-WESTKQRIARLENERLKHDLERSQTTFGRITW-----TTSQELDRAQRADK 322

QY       63 -----RGSNVTRGIINDPQRAK-----EALDKYLENHLDLKTNEGLKTENGSLGATE 109
          ||| :|| :||| - - - - - :||| :||| :|||
Db       323 ASAEILRTQAELRVFTOSDAAREFAAALQEKLSESGEVYVRKAKLENAGQEQESLRQE 382

```

Search completed: May 10, 2001, 08:39:39
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:08 ; Search time 31.03 Seconds
(without alignments)
84.257 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 1954
Sequence: 1 ACMVATRSQDTLEKVOERA.....KNDKLSLSKAKIQLEARK 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 39611 seqs, 6826361 residues

Total number of hits satisfying chosen parameters: 39611

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/pct-US01-04098A-3510.ppt.*
2: /cgn2_6/ptodata/2/paa/pct-US01-04098A-3188.ppt.*
3: /cgn2_6/ptodata/2/paa/pct-US01-04098A-1800.ppt.*
4: /cgn2_6/ptodata/2/paa/pct-US01-04098A-3768.ppt.*
5: /cgn2_6/ptodata/2/paa/pct-US01-04098A-1220.ppt.*
6: /cgn2_6/ptodata/2/paa/pct-US01-04098A-3914.ppt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	256	13.1	440	4	US-08-302-756C-35
2	207.5	10.6	931	1	PCT-US01-04098A-3150
3	207.5	10.6	990	1	PCT-US01-04098A-1182
4	184	9.4	1879	1	PCT-US00-35017A-1265
5	177.5	9.1	1047	1	PCT-US01-04098A-1169
6	177.5	9.1	1047	1	PCT-US01-04098A-3137
7	176	9.0	2816	1	PCT-US01-10484-145
8	175	9.0	1960	1	PCT-US01-04098A-1516
9	174	8.9	1963	1	PCT-US01-04098A-3484
10	173	8.9	1203	1	PCT-US01-04098A-1926
11	170.5	8.7	484	1	PCT-US01-04098A-1647
12	170.5	8.7	533	1	PCT-US01-04098A-3615
13	165	8.4	902	6	US-60-248-505-973
14	163.5	8.4	992	6	US-60-248-505-1112
15	155	7.9	456	1	PCT-US01-10484-76
16	152	7.8	502	1	PCT-US01-03782A-1066
17	151.5	7.8	954	1	PCT-US00-35017A-1117
18	151.5	7.8	1193	1	PCT-US00-35017A-1117
19	151	7.7	1002	5	US-09-604-287A-475
20	151	7.7	1002	5	US-09-604-287A-475
21	151	7.7	1095	5	US-09-834-759-493
22	146.5	7.5	432	5	US-09-834-759-493
23	146.5	7.5	432	5	US-09-825-301-18
24	146.5	7.5	432	5	US-09-604-287A-181
25	145.5	7.4	1275	1	PCT-US01-04098A-1583
26	145.5	7.4	1301	1	PCT-US01-04098A-3551
27	143.5	7.3	963	1	PCT-US01-04098A-1542

28	143.5	7.3	979	1	PCT-US01-04098A-3510
29	140.5	7.2	1160	1	PCT-US01-04098A-3188
30	140	7.2	612	1	PCT-US01-04098A-1800
31	140	7.2	690	1	PCT-US01-04098A-3768
32	139.5	7.1	1179	1	PCT-US01-04098A-1220
33	139	7.1	2515	1	PCT-US01-04098A-3914
34	139	7.1	2515	1	PCT-US01-04098A-3915
35	138.5	7.1	445	5	US-09-825-301-29
36	138.5	7.1	445	5	US-09-604-287A-473
37	138.5	7.1	445	5	US-09-834-759-473
38	135	6.9	2128	6	US-60-248-505-927
39	133.5	6.8	855	1	PCT-US01-04098A-1128
40	133.5	6.8	873	1	PCT-US01-04098A-3096
41	133	6.8	977	1	PCT-US01-04098A-1201
42	133	6.8	1003	1	PCT-US01-04098A-3169
43	133	6.8	1053	5	US-09-724-519-2
44	132.5	6.8	779	1	PCT-US00-35017A-856
45	131.5	6.7	710	1	PCT-US01-04098A-1159

ALIGNMENTS

RESULT 1
US-08-302-756C-35
; Sequence 35, Application US/08302756C
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; FILE REFERENCE: THE SURFACE OF GRAM POSITIVE BACTERIA
; CURRENT APPLICATION NUMBER: 016921-076
; CURRENT FILING DATE: 1995-03-07
; PRIOR APPLICATION NUMBER: US/08/302,756C
; PRIOR FILING DATE: 1990-05-11
; PRIOR APPLICATION NUMBER: US 07/742,199
; PRIOR FILING DATE: 1991-08-05
; PRIOR APPLICATION NUMBER: US 07/814,823
; PRIOR FILING DATE: 1991-12-23
; PRIOR APPLICATION NUMBER: US 07/851,082
; PRIOR FILING DATE: 1992-03-13
; PRIOR APPLICATION NUMBER: PCT/US93/02355
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-08-302-756C-35

Query Match	13.1%	Score 256:	DB 4:	Length 440:
Best Local Similarity	30.8%	Pred. No. 1.8e-09:		
Matches 100:	Conservative	34:	Mismatches 111:	Indels 80:
Gaps	12:			
QY	126	RVFPGTGVNPDKARELLNKYDVNSMLQANNDKLPVRVRYTRHTPDGLKLTIDDLDAK	185	
DB	1	RVFPGTGVNPDKARELLNKYDVNSMLQANNDKLT-----TENNLTQDNKLTTE	52	
QY	186	EHELOQQNEKLSLQNGDNPREVIEDLAANNPAI--QNTLRHENKDLKARLENAMVAG	243	
DB	53	KNKLTQDNKLTTE--KNLTDQNKLTTEKELKAGNRLTTEKGLTKLSEAEAAA	110	
QY	244	---RDFKRA--GTLDDQVYQLYTHNSNYQYNAQAGRL-----DLRQKAEYKLGUN	290	
DB	111	NKERENKAIGTKLTKLTDVTKDKIAKEQSKETIGTLAKTIDETVKDKIAKEQSKETI	170	
QY	291	DWABRLLOELNIDVATRSQD-----TLEKV-----OFRADK	322	
DB	171	GTLLKTLDETVKDKIAKEQSKETIGTLAKTIDETVKDKIAKEQSKQDQKALGKOLAKK	230	

QY	308	SOTDTEK-----VOERADKFEIENNTLKN-----SDLSFNNKALKOINDE	350
Db	567	KENEILQKKITNLKITEKTEALEQSELERNNKLTDLDSFNNLTQLESKEKNSQ	626
QY	351	LTEELSNAKEKLKNDKSL-----SEKASKIQ-----EUEARK	383
Db	627	LDEE-----NLELRNVESLCKASMKMAQLNKENLESEK	662

RESULT

```

RESUME
PCT-US01-04098A-1182
SEQUENCE 1182, Application PC/TUSO104098A
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: Novel Nucleic Acids
FILE REFERENCE: 21272-029
CURRENT APPLICATION NUMBER: PCT/US01-04098A
CURRENT FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: Not Yet Assigned
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ IDS NOS: 3960
SOFTWARE: Custom
SEQ ID NO 1182
LENGTH: 990
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-04098A-1182

```

Query Match	10.6%	Score	207.5;	DB 1;	Length	990;
Best Local Similarity	23.08;	Pred. No.	4.7e-06;			
Matches 106;	Conservative	86;	Mismatches	151;	Indels	117

QY	7	RGSTDTLEKVOERADKTEIENNTLKLKNSDLGNFN---	KALKDHNDLTELTELSSNAKEKUR	63
Db	296	ROELDALREKAVRVDKLESEVRSYKERLHDIEFYKARVEELKEDNQVLLETKTLMLEDOL	355	
QY	64	GSVARTGTINDPQAK--EALDKVYLENHDLKTKNEGLKTENEGLKTENEGLKTENEGL---	120	
Db	356	GT-----RARSDKLEHEKENLQKLAKLHDMERMEDMRKKTEELMEENMTLEW	404	
QY	121	--KTEVDVRVFPBG-----TVENPKA-----REL-----LNKYDVNSLMQANND	158	
Db	405	AQQSMDLSHLGWLEQIQRSTSELSAPOKSLGHEVNEELTSSRLLLKLEMENOSLTKTVE	464	
QY	159	KLPRVRYTRHTPEDKLKII-----DLDLAKHELOQQNEKLSLQNGDGNPREV	208	
Db	465	EL-----RTTVDSEGNASKILKMEKENORLSKKVEILENEIVQ---EKOSLQNCNLKSDL	518	
QY	209	IEDLAANPAIQNIR-----LRHENK-----DLKARLENAMVACGRDKFRAGTL	252	
Db	519	MKEKAQLEXTIETLRENSBROKILEGENEHLNQTVSSLURQSQISAEARVYDLEKENKI	578	
QY	253	LQDVTQLYTKHNSNYOQYNAOAGRLD-----LRQKAEYLLKGLNDWAERLLQJLQELINATVR	307	
Db	579	L-----HESIKETSSKLSLTFEPKROIKELEHYKEKGERAELEENLH-----HLE	625	

[illegible]

RESULT

```

1  RESULT
2  PCT-US01-04098A-3150
3  ; Sequence 3150, Application PC/TUS0104098A
4  ; GENERAL INFORMATION:
5  ; APPLICANT: Hyseq, Inc.
6  ; TITLE OF INVENTION: Novel Nucleic Acids and
7  ; FILE REFERENCE: 21272-029
8  ; CURRENT APPLICATION NUMBER: PCT/US01/04098A
9  ; CURRENT FILING DATE: 2001-02-05
10 ; PRIOR APPLICATION NUMBER: Not Yet Assigned
11 ; PRIOR FILING DATE: 2001-01-30
12 ; PRIOR APPLICATION NUMBER: 09/728,422
13 ; PRIOR FILING DATE: 2000-11-30
14 ; PRIOR APPLICATION NUMBER: 09/593,325
15 ; PRIOR FILING DATE: 2000-10-20
16 ; PRIOR APPLICATION NUMBER: 09/563,561
17 ; PRIOR FILING DATE: 2000-09-15
18 ; PRIOR APPLICATION NUMBER: 09/654,936
19 ; PRIOR FILING DATE: 2000-09-01
20 ; PRIOR APPLICATION NUMBER: 09/620,325
21 ; PRIOR FILING DATE: 2000-07-19
22 ; PRIOR APPLICATION NUMBER: 09/598,075
23 ; PRIOR FILING DATE: 2000-06-20
24 ; PRIOR APPLICATION NUMBER: 09/560,875
25 ; PRIOR FILING DATE: 2000-04-27
26 ; PRIOR APPLICATION NUMBER: 09/496,914
27 ; PRIOR FILING DATE: 2000-02-03
28 ; NUMBER OF SEQ ID NOS: 3960
29 ; SOFTWARE: Custom
30 ; SEQ ID NO 3150
31 ; LENGTH: 931
32 ; TYPE: prnt
33 ; ORGANISM: Homo sapiens
34 ; PCT-US01-04098A-3150

```

Query Match	10.68;	Score	207.5;	DB 1;	Length	931;	
Best Local Similarity	23.08;	Pred. NO.	4.3e-06;				
Matches	106;	Conservative	86;	Mismatches	15;	Indels	117

[illegible]

```

QY 308 SQDTTLEK-----VOERADKFFEIENNTLKN-----SDLSFNKALKDHND 350
Db 626 KENELLOKITTNLTKIEKIALQENSELERNRKLKTLDSFKNLTFOLESLEKNSQ 685
QY 351 LFEELSSNAKEKLNRKDKSL-----SEKASKIQ-----ELEARK 383
Db 686 LDEE-----NLELRNRVSLKCAWMKAQOLENKELESEK 721

RESULT 4
PCT-US00-35017A-1265
; Sequence 1265, Application PC/TUS0035017A
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: PCT/US00/35017A
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1265
; LENGTH: 1879
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1879)
; OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3
PCT-US00-35017A-1265

```

```

Query Match          9.4%; Score 184; DB 1; Length 1879;
Best Local Similarity 20.4%; Pred. No. 0.00028;
Matches 106; Conservative 83; Mismatches 171; Indels 160; Gaps 20;

QY      6 TRSOTDLE-KYQERADKFEIENNTVLKUKNSDLSFN-----NKALDHNDDELTEELS 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     948 SROELEKLRKLEGRDSFHEQIADLOAQIAELKQWLAKKEEELQAAALRDLDETAQK-N 1006
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     57 NAKEKURGSAVTRGTINDPQR-----AKEALDKYLENHDLTKTNEGKLTENE-----GLKT 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    1007 NALKTKIR---ELEGHISDLQEDLDSERAAARKAEKQKARDLGSELEALKTELEDTLDST 1063
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    109 ENGLKTENEGKLTVDVFRFGTIVPNDPKARELLNKY-----D 147
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    1064 QQE-LPRAKEQEVTVIKKALNEETRSHEAQVQEMROKHAQVQSLTEQLEQXKRAKANLD 1122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    148 VENSIMQANDKLPWRVR---YTRHTPEDKLLKI-----IDDLDAKEH 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1123 KKKOTLEKENTDLAGELRVLQQAQVEYEHMKLQAOVQELQSKCDGERARAELNDRVH 1182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    188 ELQOQNEKLS--LONGDGNPREVTEDLAANNPAIGN-----RLRH 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1183 KLQNEVESVTGMLNEAGKAIKLAKDVASLSSOLOQDTQELLQEESRQKLNVSTSLRQLEE 1242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    227 ENKDKLARLENAMVAG-----RDFKAGTLLDQVTVQLYTKKNSN 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1243 ERNSLQDQDDEMEAKQNLHERHISTLNIQDSKKKLQDFSTVEALEEGKRFQKIEN 1302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    267 Y-QOYNQAQGLDLDROKAEYKLGINDWAERLLQELN-----IDVATRSQ-TDTRLEKVVQBRAD 321
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1303 LTQQTTEKAAAYDKLETK-----NRLQOELDLVDLDNQROLVSNLEKXKQKFD 1353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    322 KFEIENNTKLKNSD-----LSFNKALKDHP-- 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   1354 QLLAEKNISSKYADERDRVEAREKETKALSRLARALEEALEAKEELERTNKMLKLAEMG 1413
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   348 -----NDELTELSNAKEKLRK--NDKSLSEKASKIOEL 379
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      1414  RFGSASKDVQGEIASHLSEKSKRALGDPRLPEEMKTOLEEL 1453

RESULT      5
PCT-US01-04098A-1169
; Sequence 1169, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1169
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1169

Query Match      9.1%; Score 177.5; DB 1; Length
Best Local Similarity 22.3%; Pred. No. 0.00034;
Matches 94; Conservative 62; Mismatches 149; Indels 1

QY 13 LEKVQRADKFEIENNTIKLNSDLSFNKAL-----KQHND-----E
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 497 LTTVAEKSRLVQLLEE-LTLRGEIEELQOCLLHSGPPPPHPDRAETILR
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 60 EKLRGSATVRCGTINDPORAKEALDKYLENHDLKTKNEGLKTENEGLKTE
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 556 EHORESGVLR-----DKYEKALKAYQAEVDKLRANAEKYAQEAVAGLKDV
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 120 LKTEVDYRFRGTGVENPDKARELLNKYDVNSMLQANDKLPWVRVTRHT
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 610 LM-----
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 180 DDLDAKEHEIQOQNEKLSLQNGDGNPREVIEDLAANNPAIONILRHE--
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 619 DSL-ASDHQKSLDKLTKATLNSGPAQQKEIGELKA---VMGIKMEHOLEL
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 236 ENAMEVAGDRFKR-----AGTLIDQVTL---YTKHNSNYOQYNAQ
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 675 ETAMHYKEKALEKLEQAEELAGLRHWAQLEVQASQHRLELQEAQDQ
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 281 -----QKAEYILKGLNDWAERLLQELNIDVATRSOTDTLEKVOE
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 735 ELEKLDVEYRQGAQALEFLKEQLSLAEKKM-LDYERLQRAEAQGQEVES
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 328 NTLKLNSDLSFNKALDKHNDTEELTESNAKE-----KLKRNKDSLSE
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 793 NRQAVALECCSOHTHTMESND-ISETTIRTKETVEGLQDKLNKRDKVTA
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |
QY 381 AR 382
| | | | | : : | : : | : : | : : | : : | : : | : : | : : |

```

[illegible]


```

; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1926
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1926

```

```

Query Match      8.9%; Score 173; DB 1; Length 1203;
Best Local Similarity 21.4%; Pred. No. 0.00077;
Matches 89; Conservative 61; Mismatches 142; Indels 124; Gaps 13;

```

```

QY 9 QDTTLEKQVE-----RADKFEI-----NNTLKNKNSDLSPFNKALKDHNDELTE 53
DB 815 EQOTLNRALEEGKOREVLRGRGAELDEQKRLDRTVDRLNKELEKIGESKQALQLOA 874
QY 54 ELSNAKEKLRSV-----TRGTI-----NDPQRAKEALDKYELNHDLT 94
DB 875 QLEDYKEKARREVDARQAKDWASAEKTSGLSLQDEIQRQLAQASQAEERTARL 934
QY 95 KNEGLKTENEGKLTENEGKLTENEGKLTENEGKLTENEGKLTENEGKLTENEGKLT 154
DB 935 DRELLAQRLQGLEOEAENKRSOD-----DRARQLG-----LE 968
QY 155 ANNDKLPVRVRYTRHTPE-----DKLKKTIIDDLQAKEHELOQNEKLSLQNGDGNPREVIED 211
DB 969 EKYSRLETDELDEKNTVELLTDVRNRCQVDQRLTELMOE-----RSARQD 1015
QY 212 LAANNPAIONTELREHKNKDLKARLENAMVAGNDFKAGTLLDQVTLQYTKHNSYQOYN 271
DB 1016 LECD-----KISLRQNKDLTRLAS-----SEGOKPSASLSQLE 1051
QY 272 AQGRILDRQKAE-----YKGLNDWAERLLQELNIDVATRSQTDITLEKYOEADKFEI 326
DB 1052 SQNLQOERLQAEEREKTVLQSTNRKLRKVKELSIIQI-----EDERQHVNDQKQ----- 1102
QY 327 NNTLKNKNSDLSPFNKALKDHNDELTELSNAKEKLKRNKDSLSKASKIOLEAR 382
DB 1103 -----LSLRVYKALKRQVDEAEETELRDLGRLKKAQREVEEQHEVNEQLQAR 1148

```

```

RESULT 11
PCT-US01-04098A-1647
; Sequence 1647, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20

```

```

; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1647
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1647

```

```

Query Match      8.7%; Score 170.5; DB 1; Length 484;
Best Local Similarity 21.6%; Pred. No. 0.00035;
Matches 91; Conservative 73; Mismatches 134; Indels 123; Gaps 16;

```

```

QY 12 TLEKYOEADKFEIENNTLKNKNSDLSPFNKALKDHNDELTELSNAKEKLRSVATRG 71
DB 92 TIERLKEQORDEREK-----QEEIDNVYKDKLKEKV-----SLQGD 131
QY 72 INDPQRAKEALDKYELNHDLTKNNEGLKTENEGKLT-----ENEGKLTENEGKLT 123
DB 132 LSE-----KEA-SLLDLKEHASSLSSGLKKDSR-LKLTLEIALEQKKEELKMSQLKAH 185
QY 124 VDRVFPRTGVENPKAREL-----LNKYDVNSMLQANDKLPWRYTRTPEDKLLKIID 180
DB 186 EAALERARSPMSDRIQHLEREITRYKDESSAKAQVEDRLEILKEVENKDKDKKIAE 245
QY 181 DUDAKEHELOQNEKLSLQNGDGNPREVIEDLAANNPAIONIRLREHKNKDLKARLENAME 240
DB 246 LESLTSROVKDKNKVA-----NLKHKEQVEKKKSAQMLE 280
QY 241 VAGRDFKAGTLLDQVTLQYTKHNSYQOYNQAQGRILDRQKAEYKGLND-----W 292
DB 281 EARR-----REDNLNDSQLOD-----SLRKKDDRIEELAEALRESVQIT 321
QY 293 AER-----LLOELNIDVATRSQTDITLEKYOEADKFEIENNTLKNKNSDLSPFNKAL----- 344
DB 322 AERWVLAQESARTNAEKQVEELLMAEKV-KQELSWKAKLSSTOOSLAKEKETHLTWL 380
QY 345 -----KQHNDELTE-----ELSNNAKEKLKRNKDSLS-ERAKSIQEL 379
DB 381 RAERKHLVEYLEMKQEPALLAATSEKDNALIELLSSSKKTKTOEVAALKREKDRLVQOL 440
QY 380 E 380
DB 441 K 441

```

```

RESULT 12
PCT-US01-04098A-3615
; Sequence 3615, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20

```


QY	64	-----GSAVTRGTINDPQRAKALDKYELNHDLTCKNEGLKTENEGLKTKTEN 117	Db	133	PNFLQDSVKSLT-----RSKEELSR-----VKCRNAALK-ESQKLKEDLEAVEDR---E 180
Db	396	YKVGAGSSLYGGTITINARKFEEMNAELENKELA---QNRLECELEKLRQDFEETTON 452	QY	135	NPKAR--ELLNKYDVENSMLQANNDKLPWRVRYTRHTPEDKLLKIIDDLDAKEHELOQ 192
QY	118	EGLTEVDVPRPTVNP-----DKARELL-----NKYDV 148	Db	181	N-KKARNOSLLTVGNFORQLAEAKEDNC-----KVTIMLENVLASHSKMOGA 226
Db	453	EKLVELRSAVEQVVKETPEYRCMQSFVLYNESLQKKAHLDEARTLLTGTHQHV 512	QY	193	NEKLSLONG--DGNPREVIEDLAANPAIQNIRLR-----HENKDLKARLEN 237
QY	149	ENSMLOANNDKLPWRVRYTRHTPEDKLLKIIDDLDAKEHELOQ---QNEKL----- 196	Db	227	LEKVQIELGRRRSEIAGLKKERDLNQORVQKLEAEVDQWQARMVMEQHNSEIES-LQK 285
Db	513	E--LIERDEVSLLKRLTEVTLQEDTLAQYRKEVEMLRIEFQOTLAANEQAQPINREMRH 570	QY	238	AMEVAGRDCKFRAGTLLDQVQTYLTKHNSYQYNAQAGRLDLRQKAEYLGKLNDAERLL 297
QY	197	---SLQNGD-----GNPREVIEDL-----AANPAIQNIRLRH 226	Db	286	ALGVAREDNKRKLAMSLQALQ-----TNNHLQTKLDHIQEQLESKELEPERONLETCKDRMT 340
Db	571	LISSLOHNHQLGVEVLYRKKLRQAQSDLNKTRLSGSSALLOSQSTEDPKDEPAELKP 630	QY	298	QELNIDVATRSQTDITLEKVQERADKFEIENNTLK-----LKNSDLSF 339
QY	227	ENKOL-----KARLENAMEV--AGRDFKFRAGTLLDQVQTYLTKHNSYQ---QYNAQ 273	Db	341	ESKVEAELHA-----ERTEALRKQFQTERETTKKVAQREVAELKKALDEANFRSVEYSR 395
Db	631	DSEDLSSOSSASKASQEDANEIKKAQESQKEMKLLLDWYRSAPKEQDKYQLMAAEKKS 690	QY	340	NKALKADHNDLTELTELSNAKEKLRKNDKS-----LSEKASKIOELEARK 383
QY	274	AGRLDLRQKAEYLGKLNDAERLLQELNIDVATRSQTDITLEKVQERADKFEIENNTLK 333	Db	396	TNRELQKLAELKELKILESNEKI-KNOKTQIKLHLSAKANNAQNIERMK 443
Db	691	AELEDLRQR---LKDLEDEKKEKNKM-----ADEDALRKIRAVEEQIEYLGKKLAMA 740			
QY	334	N-----SDLSFNKALKDHNDLTELTELSNAKEKLRKNDKSLSEKASKIOEL 379			
Db	741	KOEERALLSEMDVTGQAFDMQEQNIRLMQOLREKDDANFKMISQINQIHLK 793			

Search completed: May 10, 2001, 08:38:50
Job time: 102 sec

RESULT 15
PCT-US01-10484-76
; Sequence 76, Application PC/TUS0110484
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-048
; CURRENT APPLICATION NUMBER: PCT/US01/10484
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/668,680
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/695,618
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/728,711
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: Custom
; SEQ ID NO 76
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-10484-76

Query Match 7.9%; Score 155; DB 1; Length 456;
Best Local Similarity 20.7%; Pred. No. 0.0029;
Matches 97; Conservative 81; Mismatches 159; Indels 132; Gaps 20;

QY	5	ATRSQTDITLEKVQERADKFEIENNTLKNSDLSFNKALKDHNDLTELTELS----- 56
Db	17	ATASETFV--LQORMRIVEEQTSURDDLIMLDFCEK--RSHVNSIKERTSSVGLPSVI 72
QY	57	-----NAKEKLGRS-----AVTRGTIND-- 74
Db	73	PNSTRRVSPAPNLPMSKTSQDIGDSRISLTKLLNAIKTMEGRLEGKIEILASRPLINDES 132
QY	75	PQRAKALDKYELNHDLTCKNEGLKTENEGLKTEVEVDVPRPTVTE 134

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:40:13 ; Search time 44.17 Seconds
(without alignments)
1394.777 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 383

Sequence: 1 ACWVATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1009251 seqs, 160854530 residues

Word size: 10

Total number of hits satisfying chosen parameters: 24

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: Pending Patents_AA_Main:*

```

1: /cgn2_6/ptodata/2/paa/pctus_comb.pep.*
2: /cgn2_6/ptodata/2/paa/us06_comb.pep.*
3: /cgn2_6/ptodata/2/paa/us07_comb.pep.*
4: /cgn2_6/ptodata/2/paa/us08_comb.pep.*
5: /cgn2_6/ptodata/2/paa/us081_comb.pep.*
6: /cgn2_6/ptodata/2/paa/us082_comb.pep.*
7: /cgn2_6/ptodata/2/paa/us083_comb.pep.*
8: /cgn2_6/ptodata/2/paa/us084_comb.pep.*
9: /cgn2_6/ptodata/2/paa/us085_comb.pep.*
10: /cgn2_6/ptodata/2/paa/us086_comb.pep.*
11: /cgn2_6/ptodata/2/paa/us087_comb.pep.*
12: /cgn2_6/ptodata/2/paa/us088_comb.pep.*
13: /cgn2_6/ptodata/2/paa/us089_comb.pep.*
14: /cgn2_6/ptodata/2/paa/us090_comb.pep.*
15: /cgn2_6/ptodata/2/paa/us091_comb.pep.*
16: /cgn2_6/ptodata/2/paa/us092_comb.pep.*
17: /cgn2_6/ptodata/2/paa/us093_comb.pep.*
18: /cgn2_6/ptodata/2/paa/us094_comb.pep.*
19: /cgn2_6/ptodata/2/paa/us095_comb.pep.*
20: /cgn2_6/ptodata/2/paa/us096_comb.pep.*
21: /cgn2_6/ptodata/2/paa/us097_comb.pep.*
22: /cgn2_6/ptodata/2/paa/us098_comb.pep.*
23: /cgn2_6/ptodata/2/paa/us060_comb.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	383	100.0	383	15	US-09-151-409-16
2	50	13.1	443	7	US-08-325-278-6
3	35	9.1	440	17	US-09-302-756-35
4	33	8.6	42	1	PCT-US94-12268-31
5	33	8.6	42	5	US-08-143-412A-31
6	33	8.6	42	8	US-08-479-690-31
7	27	7.0	28	5	US-08-187-166-20
8	16	4.2	254	13	US-08-914-479-4
9	15	3.9	15	3	US-07-945-860B-4
10	15	3.9	15	3	US-07-945-860B-5

```

11 15 3.9 15 3 US-07-945-860B-6
12 15 3.9 15 3 US-07-945-860B-7
13 15 3.9 15 3 US-07-945-860B-9
14 15 3.9 15 13 US-08-914-479-7
15 15 3.9 15 13 US-08-914-479-8
16 15 3.9 15 13 US-08-914-479-9
17 15 3.9 15 13 US-08-914-479-10
18 15 3.9 15 13 US-08-914-479-12
19 15 3.9 281 13 US-08-914-479-6
20 14 3.7 14 5 US-08-187-166-15
21 13 3.4 13 5 US-08-187-166-23
22 12 3.1 138 13 US-08-914-479-2
23 11 2.9 12 5 US-08-187-166-16
24 10 2.6 10 5 US-08-187-166-24

```

ALIGNMENTS

```

RESULT 1
US-09-151-409-16
; Sequence 16, Application US/09151409A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: GROUP A STREPTOCOCCAL VACCINES
; FILE REFERENCE: 48112.410
; CURRENT APPLICATION NUMBER: US/09/151.409A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,635
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Hexavalent M
; OTHER INFORMATION: fusion gene sequence constructed from
; OTHER INFORMATION: streptococcal type 24, 5, 6, 19, 1 and 3 M protein
; OTHER INFORMATION: DNAs
US-09-151-409-16

```

```

Query Match 100.0%; Score 383; DB 15; Length 383;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACWVATRSQDTLEKVOERADKFEIENNTLKNKNSDLSPFNKALKDHNDLTELSSNAKE 60
Db 1 ACWVATRSQDTLEKVOERADKFEIENNTLKNKNSDLSPFNKALKDHNDLTELSSNAKE 60

Qy 61 KLRSVATRGTTINDPQAKALKDYELNHDLTKNKNEGLKTENEGLKTENEGL 120
Db 61 KLRSVATRGTTINDPQAKALKDYELNHDLTKNKNEGLKTENEGLKTENEGL 120

Qy 121 KTEVDVFPFGTGVENPDKARELLNKYDVNSMLQANNKLPWVRVYTRHTPDKLLKIID 180
Db 121 KTEVDVFPFGTGVENPDKARELLNKYDVNSMLQANNKLPWVRVYTRHTPDKLLKIID 180

Qy 181 DLDAKEHELOQNEKLSLQNGDGNPREVIEDLAANPAIONIRLRHENKDLKARLENAM 240
Db 181 DLDAKEHELOQNEKLSLQNGDGNPREVIEDLAANPAIONIRLRHENKDLKARLENAM 240

Qy 241 VAGDFRAGTLLDOVTLTKNSYQYNAQAGRLDLRQKAEYLKGLNDWAERLLQEL 300
Db 241 VAGDFRAGTLLDOVTLTKNSYQYNAQAGRLDLRQKAEYLKGLNDWAERLLQEL 300

Qy 301 NIDVATRSQDTLEKVOERADKFEIENNTLKNKNSDLSPFNKALKDHNDLTELSSNAKE 360
Db 301 NIDVATRSQDTLEKVOERADKFEIENNTLKNKNSDLSPFNKALKDHNDLTELSSNAKE 360

Qy 361 KLRNKDKSLSEKASKIQELEARK 383

```

```
Db 361 KLRKNDKSLSEKASKIQLEARK 383
|||||
; EARLIER APPLICATION NUMBER: PCT/US93/02355
; EARLIER FILING DATE: 1993-03-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-09-302-756-35

Query Match 9.1%; Score 35; DB 17; Length 440;
Best Local Similarity 100.0%; Pred. No. 4.4e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRTGVNPKARELLNKYDVNSMLQANNDKL 160
|||||
Db 1 RVFPRTGVNPKARELLNKYDVNSMLQANNDKL 35

RESULT 4
PCT-US94-12268-31
; Sequence 31, Application PC/TUS9412268
; GENERAL INFORMATION:
; APPLICANT: United Biomedical, Inc.
; TITLE OF INVENTION: Structured Synthetic Antigen
; TITLE OF INVENTION: Libraries (SSAL) for Diagnostics, Vaccines and
; TITLE OF INVENTION: Therapeutics
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin, Esq.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12268
; FILING DATE: 26-OCT-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/143,412
; FILING DATE: 26-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin, Maria C.H.
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4120PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 2
; OTHER INFORMATION: /note= "T1.Vi;*"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "V1.F1;*"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note= "T1.N1.P1"
```

```
Db 361 KLRKNDKSLSEKASKIQLEARK 383
|||||
; EARLIER APPLICATION NUMBER: PCT/US93/02355
; EARLIER FILING DATE: 1993-03-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-09-302-756-35

Query Match 13.1%; Score 50; DB 7; Length 443;
Best Local Similarity 100.0%; Pred. No. 4.7e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANPAIQIRLRHENKDLKARLENAMVAGRDFKRA 249
|||||
Db 1 NGDGNPREVIEDLAANPAIQIRLRHENKDLKARLENAMVAGRDFKRA 50

RESULT 3
US-09-302-756-35
; Sequence 35, Application US/09302756
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; TITLE OF INVENTION: THE SURFACE OF GRAM POSITIVE BACTERIA
; FILE REFERENCE: 016921-076
; CURRENT APPLICATION NUMBER: US/09/302,756
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: US 07/522,440
; EARLIER FILING DATE: 1990-05-11
; EARLIER APPLICATION NUMBER: US 07/742,199
; EARLIER FILING DATE: 1991-08-05
; EARLIER APPLICATION NUMBER: US 07/814,823
; EARLIER FILING DATE: 1991-12-23
; EARLIER APPLICATION NUMBER: US 07/851,082
; EARLIER FILING DATE: 1992-03-13
```


/	NAME/KEY:	Modified-site	
/	LOCATION:	23	
/	OTHER INFORMATION:	/note= "L1;T1;V1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	24	
/	OTHER INFORMATION:	/note= "E2;Q1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	26	
/	OTHER INFORMATION:	/note= "H1;I1;S1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	27	
/	OTHER INFORMATION:	/note= "D1;R1;M1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	29	
/	OTHER INFORMATION:	/note= "K1;R1;Q1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	30	
/	OTHER INFORMATION:	/note= "T1;H1;A1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	31	
/	OTHER INFORMATION:	/note= "K1;E1;N1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	33	
/	OTHER INFORMATION:	/note= "E1;K1;D1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	34	
/	OTHER INFORMATION:	/note= "G1;D1;K1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	36	
/	OTHER INFORMATION:	/note= "K2;T1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	37	
/	OTHER INFORMATION:	/note= "T2;A1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	38	
/	OTHER INFORMATION:	/note= "E2;R1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	39	
/	OTHER INFORMATION:	/note= "N2;L1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	40	
/	OTHER INFORMATION:	/note= "E2;N1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	41	
/	OTHER INFORMATION:	/note= "G1;N1;L1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	42	
/	OTHER INFORMATION:	/note= "L2;A1"	
/	FEATURE:		
/	NAME/KEY:	Modified-site	
/	LOCATION:	43	
/	OTHER INFORMATION:	/note= "PCT-USA94-122668-31"	

QY	74	DPQRAKALDKYELENHDLKTKNEGLKTENEGL	106
Dd	10	DPQRAKALDKYELENHDLKTKNEGLKTENEGL	42

RESULT 5
US-08-143-412A-31
; Sequence 31, Application US/08143412A
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy J.
; APPLICANT: Ye, John
; APPLICANT: Kaminsky, Stephen M.
; APPLICANT: Hosein, Barbara H.
; APPLICANT: Nixon, Douglas F.
; APPLICANT: Koif, Wayne C.
; APPLICANT: Kowalski, Jacek
; TITLE OF INVENTION: Structured Synthetic Antigen Libraries
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,412A
; FILING DATE: 25-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 2008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-2828 x2119
; TELEFAX: 516-273-1717
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 2
; OTHER INFORMATION: /note= "T1;V1;*"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "V1;F1;*"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note= "T1;N1;P1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5
; OTHER INFORMATION: /note= "R2;G1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 6
; OTHER INFORMATION: /note= "G2;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note= "T2;G1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 8

; OTHER INFORMATION: /note= "I1;N1;V1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: /note= "S1;P1;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "D1;R1;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 11
; OTHER INFORMATION: /note= "P2;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Q1;V1;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /note= "R1;I1;K1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "A2;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "K1;D1;R1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 16
; OTHER INFORMATION: /note= "E2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 17
; OTHER INFORMATION: /note= "A2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "L2;A1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 19
; OTHER INFORMATION: /note= "D1;N2"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 20
; OTHER INFORMATION: /note= "K2;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 21
; OTHER INFORMATION: /note= "Y2;P1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 22
; OTHER INFORMATION: /note= "E1;A1;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "L1;I1;V1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 24
; OTHER INFORMATION: /note= "E2;Q1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "H1;I1;S1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "D1;R1;M1"

FEATURE:
NAME/KEY: Modified-site
LOCATION: 29
OTHER INFORMATION: /note= "K1;R1;Q1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 30
OTHER INFORMATION: /note= "T1;H1;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 31
OTHER INFORMATION: /note= "K1;E1;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 33
OTHER INFORMATION: /note= "E1;K1;D1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 34
OTHER INFORMATION: /note= "G1;D1;K1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 36
OTHER INFORMATION: /note= "K2;T1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 37
OTHER INFORMATION: /note= "T2;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 38
OTHER INFORMATION: /note= "E2;R1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 39
OTHER INFORMATION: /note= "N2;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 40
OTHER INFORMATION: /note= "E2;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 41
OTHER INFORMATION: /note= "G1;N1;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 42
OTHER INFORMATION: /note= "L2;A1"
US-08-143-412A-31

Query Match 8.6%; Score 33; DB 5; Length 42;
Best Local Similarity 100.0%; Pred. No. 4e-25;
Matches 33; ~ Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 DPQRAKALDYELNHDLTKEGLKTENGL 106
|||||
Db 10 DPQRAKALDYELNHDLTKEGLKTENGL 42

RESULT 6
US-08-479-690-31
Sequence 31, Application US/08479690
GENERAL INFORMATION:
APPLICANT: United Biomedical, Inc.
TITLE OF INVENTION: Structured Synthetic Antigen Libraries
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin, Esq.
STREET: 345 Park Avenue
CITY: New York
STATE: New York

COUNTRY: U.S.A.
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,690
FILING DATE: 7-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/143,412
FILING DATE: 26-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lin, Maria C.H.
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4120US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /note= "T1;V1;*"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note= "V1;F1;*"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "T1;N1;P1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 5
OTHER INFORMATION: /note= "R2;G1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /note= "G2;D1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7
OTHER INFORMATION: /note= "T2;G1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 8
OTHER INFORMATION: /note= "T1;N1;V1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /note= "S1;P1;E1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 10
OTHER INFORMATION: /note= "D1;R1;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 11
OTHER INFORMATION: /note= "P2;E1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 12
OTHER INFORMATION: /note= "Q1;V1;D1"
NAME/KEY: Modified-site

Tue May 15 07:28:35 2001

```
LOCATION: 13
OTHER INFORMATION: /note= "R1;I1;K1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 14
OTHER INFORMATION: /note= "A2;E1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /note= "K1;D1;R1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 16
OTHER INFORMATION: /note= "E2;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 17
OTHER INFORMATION: /note= "A2;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 18
OTHER INFORMATION: /note= "L2;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 19
OTHER INFORMATION: /note= "D1;N2"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 20
OTHER INFORMATION: /note= "K2;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 21
OTHER INFORMATION: /note= "Y2;P1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 22
OTHER INFORMATION: /note= "E1;A1;D1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 23
OTHER INFORMATION: /note= "L1;I1;V1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 24
OTHER INFORMATION: /note= "E2;Q1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 26
OTHER INFORMATION: /note= "H1;I1;S1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 27
OTHER INFORMATION: /note= "D1;R1;M1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 29
OTHER INFORMATION: /note= "K1;R1;Q1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 30
OTHER INFORMATION: /note= "T1;H1;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 31
OTHER INFORMATION: /note= "K1;E1;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 33
OTHER INFORMATION: /note= "E1;K1;D1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 34

OTHER INFORMATION: /note= "G1;D1;K1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 36
OTHER INFORMATION: /note= "K2;T1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 37
OTHER INFORMATION: /note= "T2;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 38
OTHER INFORMATION: /note= "E2;R1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 39
OTHER INFORMATION: /note= "N2;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 40
OTHER INFORMATION: /note= "E2;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 41
OTHER INFORMATION: /note= "G1;N1;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 42
OTHER INFORMATION: /note= "L2;A1"
US-08-479-690-31

Query Match 8.6% Score 33; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 4e-25;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 DPQRAKALDKYELENHDLTKNEGLKTENEGL 106
| | | | | | | | | | | | | | | | | | | | | |
Db 10 DPQRAKALDKYELENHDLTKNEGLKTENEGL 42

RESULT 7
US-08-187-166-20
; Sequence 20, Application US/08187166
; GENERAL INFORMATION:
; APPLICANT: Klemm, Per
; APPLICANT: Sokurenko, Evgeni Veniaminovic
; APPLICANT: Hasty, David Long
; APPLICANT: Pallesen, Lars
; APPLICANT: Molin, Soren
; TITLE OF INVENTION: RECEPTOR SPECIFIC BACTERIAL ADHESINS AND
; TITLE OF INVENTION: THEIR USE
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,166
; FILING DATE: 27-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: KLEMM-1
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-187-166-20

Query Match 7.0%; Score 27; DB 5; Length 28;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 KTKNEGLKTENEGKTENEGKTENEG 119
|||||
Db 1 KTKNEGLKTENEGKTENEGKTENEG 27

RESULT 8
US-08-914-479-4
Sequence 4, Application US/08914479
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.4042
CURRENT APPLICATION NUMBER: US/08/914,479
CURRENT FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

Query Match 4.2%; Score 16; DB 13; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 MVATRSQTDLEKVOE 18
|||||
Db 1 MVATRSQTDLEKVOE 16

RESULT 9
US-07-945-860B-4
Sequence 4, Application US/07945860B
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.4042
CURRENT APPLICATION NUMBER: US/07/945,860
CURRENT FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,860B
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5770P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8394
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-945-860B-4

Query Match 3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VATRSQTDLEKVOE 18
|||||
Db 1 VATRSQTDLEKVOE 15

RESULT 10
US-07-945-860B-5
Sequence 5, Application US/07945860B
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.4042
CURRENT APPLICATION NUMBER: US/07/945,860
CURRENT FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,860B
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5770P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8394
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-945-860B-5

us-09-151-409-16.oligo.rapm

Tue May 15 07:28:35 2001

```

Query Match      3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 AVTRGTINDPQRAKE 80
DB 1 AVTRGTINDPQRAKE 15

RESULT 11
US-07-945-860B-6
; Sequence 6, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-945-860B-7

Query Match      3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RRVYTRHTPDKLLK 177
DB 1 RRVYTRHTPDKLLK 15

RESULT 13
US-07-945-860B-9
; Sequence 9, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-945-860B-9

Query Match      3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVPENPKAR 140
DB 1 RVFPRGTVPENPKAR 15

RESULT 12
US-07-945-860B-7
; Sequence 7, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.

```

Query Match 3.9%; Score 15; DB 3; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAA 214
 |||||
 Db 1 NGDGNPREVIEDLAA 15

RESULT 14
 US-08-914-479-7
 ; Sequence 7, Application US/08914479
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James B.
 ; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
 ; FILE REFERENCE: 481112.404C2
 ; CURRENT APPLICATION NUMBER: US/08/914,479
 ; PRIOR FILING DATE: 1997-08-19
 ; PRIOR APPLICATION NUMBER: 08/409,270
 ; PRIOR FILING DATE: 1995-03-23
 ; PRIOR APPLICATION NUMBER: 07/945,860
 ; PRIOR FILING DATE: 1992-09-16
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 15
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: NH2-terminal fragment of M protein for
 ; OTHER INFORMATION: constructing antigens, which elicit opsonic
 ; OTHER INFORMATION: antibodies in an immunized animal
 US-08-914-479-7

Query Match 3.9%; Score 15; DB 13; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VATRSQDTLEKQVE 18
 |||||
 Db 1 VATRSQDTLEKQVE 15

RESULT 15
 US-08-914-479-8
 ; Sequence 8, Application US/08914479
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James B.
 ; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
 ; FILE REFERENCE: 481112.404C2
 ; CURRENT APPLICATION NUMBER: US/08/914,479
 ; PRIOR FILING DATE: 1997-08-19
 ; PRIOR APPLICATION NUMBER: 08/409,270
 ; PRIOR FILING DATE: 1995-03-23
 ; PRIOR APPLICATION NUMBER: 07/945,860
 ; PRIOR FILING DATE: 1992-09-16
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 15
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: NH2-terminal fragment of M protein for
 ; OTHER INFORMATION: constructing antigens, which elicit opsonic
 ; OTHER INFORMATION: antibodies in an immunized animal
 US-08-914-479-8

Query Match 3.9%; Score 15; DB 13; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 AVTRGTINDPQRAKE 80
 |||||
 Db 1 AVTRGTINDPQRAKE 15

Search completed: May 10, 2001, 08:42:14
 Job time: 121 sec

us-09-151-409-16.oligo.rapm

Tue May 15 07:28:35 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:40:28 ; Search time 6.16 seconds
(without alignments)
424.431 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 383

Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 39611 seqs, 6826361 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	35	9.1	440 4	US-08-302-756C-35 Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-302-756C-35
; Sequence 35, Application US/08302756C
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; TITLE OF INVENTION: THE SURFACE OF GRAM POSITIVE BACTERIA
; FILE REFERENCE: 016921-076
; CURRENT APPLICATION NUMBER: US/08/302,756C
; CURRENT FILING DATE: 1995-03-07
; PRIOR APPLICATION NUMBER: US 07/522,440
; PRIOR FILING DATE: 1990-05-11
; PRIOR APPLICATION NUMBER: US 07/742,199
; PRIOR FILING DATE: 1991-08-05
; PRIOR APPLICATION NUMBER: US 07/814,823
; PRIOR FILING DATE: 1991-12-23
; PRIOR APPLICATION NUMBER: US 07/851,082
; PRIOR FILING DATE: 1992-03-13
; PRIOR APPLICATION NUMBER: PCT/US93/02355

; PRIOR FILING DATE: 1993-03-12
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-08-302-756C-35

Query Match 9.1%; Score 35; DB 4; Length 440;
Best Local Similarity 100.0%; Pred. No. 6.2e-28;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
|||||
DB 1 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 35

Search completed: May 10, 2001, 08:42:23
Job time: 115 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:38:18 ; Search time 21.4 Seconds
(without alignments)
1023.062 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 383

Sequence: 1 ACWATRSQDTLEKVVQRA.....KNDKSLSEKASKIQLEARK 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 390729 seqs, 57163235 residues

Word size : 10

Total number of hits satisfying chosen parameters: 40

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

A_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383	100.0	383	20	Streptococcal hexa
2	134	35.0	247	21	S. pyogenes hybrid
3	95	24.8	247	15	Recombinant M24-M5
4	80	20.9	287	15	Recombinant M24-M5
5	80	20.9	287	21	S. pyogenes hybrid
6	72	18.8	173	15	Recombinant M24-M5
7	72	18.8	173	21	S. pyogenes hybrid
8	72	18.8	247	15	Recombinant M19-M6
9	72	18.8	247	21	S. pyogenes hybrid
10	72	18.8	274	15	Recombinant M24-M5
11	72	18.8	274	21	S. pyogenes hybrid

12	72	18.8	343	15	R50999
13	72	18.8	343	21	B03121
14	35	9.1	100	6	P50295
15	35	9.1	441	10	P09055
16	35	9.1	441	14	R41780
17	35	9.1	483	18	W08927
18	33	8.6	42	16	R74258
19	31	8.1	31	8	P70907
20	27	7.0	28	16	R76756
21	18	4.7	67	15	R51000
22	18	4.7	67	21	B03123
23	18	4.7	135	15	R51001
24	18	4.7	135	21	B03124
25	18	4.7	187	15	R50995
26	18	4.7	187	21	B03116
27	18	4.7	305	15	R50996
28	18	4.7	305	21	B03117
29	16	4.2	234	15	R50228
30	16	4.2	284	15	R50229
31	15	3.9	15	15	R50233
32	15	3.9	15	15	R50230
33	13	3.4	13	16	R76751
34	13	3.4	13	16	R76759
35	12	3.1	13	15	R50231
36	12	3.1	138	15	R50227
37	11	2.9	12	11	R04956
38	11	2.9	12	16	R76752
39	11	2.9	13	11	R04952
40	10	2.6	10	16	R76760

ALIGNMENTS

RESULT 1

Y04368
ID Y04368 standard; Protein; 383 AA.

XX AC Y04368;

XX DT 23-JUN-1999 (first entry)

XX DE Streptococcal hexavalent M protein vaccine.

XX KW Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A; immunogenic; immune response; pathogen; fusion protein.

XX OS Streptococcus sp.

XX PN W09913084-A1.

XX PD 18-MAR-1999.

XX PF 14-SEP-1998; 98WO-US19100.

XX PR 12-SEP-1997; 97US-0058635.

XX PA (IDVA-) ID VACCINE.

XX PI Dale JB;

XX DR WPI; 1999-215066/18.

XX PT N-PSDB; X33103.

XX PS Immunogenic fusion protein derived from group A streptococci

XX CC Example 1; Fig 7; 50pp; English.

CC The present invention describes an immunogenic fusion polypeptide (I) that stimulates an immune response against group A streptococci (GAS). (I) comprises: (a) at least 2 peptides (Ia) from GAS, at least 10 amino acids long and able to induce a specific immune response; and (b) a peptide (Ib), C-terminal to (Ia) that protects the immunogenicity of the

component described in (a) but is not essential for stimulation of the immune response. Vaccines containing (I) are used to protect against GAS infections, specifically those caused by Streptococcus pyogenes, e.g. pharyngitis, pyoderma, toxic shock syndrome, deep tissue invasion, sepsis and acute rheumatic fever. (I) have improved immunogenicity and do not generate antibodies that are cross-reactive with human tissues. The present sequence represents a hexavalent M protein vaccine.

XX Sequence 383 AA;

Query Match 100.0%; Score 383; DB 20; Length 383;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACMVATRSQDTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDLTELSSNAKE 60
DB 1 acmvatrsqdtlekvoeradkfeientlknslsfnnkalkdhndelteelsnake 60
QY 61 KLKRSVTRGTINDPQRAKEALDKYLENHDLTKNEGLKTENEGKLTENEGL 120
DB 61 klkrsavtrgtindpqrakealdkyelenhdltkneglkteneglktenegkl 120
QY 121 KTEVDVFPRTGTENPKARELLNKDYVENSMLQANNDKLPWRVRYTRHTPEDKLRKIID 180
DB 121 ktevdvfpgrtvenpdkarellnkydvensmlqanndklpwrvrytrhtpedkllkiid 180
QY 181 DLDAKEHELOQNEKLSIQNGDNPREVIEDLAANPAIONIRLRHENKDKLKALENAME 240
DB 181 dldakeheiqgnneklsiqngdnprevieelaanpaignrlrhenkdkaalename 240
QY 241 VAGRDFKRAAGTLDDQVTLQYTKHNSNQYNAQAGRLDLRQKAEYKGLNDWAERLLQEL 300
DB 241 vagrdfkragtlldqvltqytkhnsnyqynaqagrldlrqkaeykglndwaerllqel 300
QY 301 NIDVATRSQDTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDLTELSSNAKE 360
DB 301 nidvatrsqdtlekvoeradkfeientlknslsfnnkalkdhndelteelsnake 360
QY 361 KLKNDKSLSEKASKIOLEARK 383
DB 361 klkndkslsekaskioleark 383

RESULT 2
B03114
ID B03114 standard; Protein; 247 AA.
XX AC B03114;
DT 10-OCT-2000 (first entry)
XX DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:4.
XX KW Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia.
XX OS Streptococcus pyogenes.
OS Synthetic.
XX PN US6063386-A.
XX PD 16-MAY-2000.
XX PF 15-SEP-1997; 97US-0937271.
XX PR 16-SEP-1992; 92US-0945954.
XX PA (UYTE-) UNIV TENNESSEE RES CORP.

PI Lederer JW, Dale JB;
XX WPI: 2000-364475/31.
DR N-PSDB; A57894.
XX
PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX
PS Disclosure: Fig 4A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B0313-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
XX invention.

XX Sequence 247 AA;

Query Match 35.0%; Score 134; DB 21; Length 247;
Best Local Similarity 100.0%; Pred. No. 3.7e-113;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 GSAVTRGTINDPQRAKEALDKYLENHDLTKNEGLKTENEGKLTENEGKLTENEGKLT 123
DB 114 gsavtrgtindpqrakealdkyelenhdltkneglkteneglkteneglkteneglk 173
QY 124 VDRVFPRTGTENPKARELLNKDYVENSMLQANNDKLPWRVRYTRHTPEDKLRKIIDLD 183
DB 174 vdrvfpgrtvenpdkarellnkydvensmlqanndklpwrvrytrhtpedkllkiidld 233
QY 184 AKEHELOQNEKLS 197
DB 234 akehelqgnnekls 247

RESULT 3
R50993
ID R50993 standard; Protein; 247 AA.
XX AC R50993;
XX DT 02-NOV-1994 (first entry)
XX DE Recombinant M24-M5-M6-M19.
XX KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

Db 2 vatrsqtdtlekvqradkfeienntklknsdlsfnkalkdhndelteelsnakeklr 61

QY 364 KNDKSLSEKASKIQELEAR 383
|||||

Db 62 kndkselsekaskiqelear 81

RESULT 5

B03113

ID B03113 standard; Protein; 287 AA.

XX

AC B03113;

XX

DT 10-OCT-2000 (first entry)

XX

DE S. pyogenes hybrid M protein (M24-M5-M6), SEQ ID NO:2.

XX

XX Multivalent hybrid M protein; group A streptococcus; serotype;
immunogenic; sero-specific antibody; streptococcal infection;
cross reactivity; vaccine; acute rheumatic fever; ARF;
rheumatic heart disease; streptococcal pharyngitis; strep throat;
pneumonia.

XX

OS Streptococcus pyogenes.

OS Synthetic.

XX

PN US6063386-A.

XX

PD 16-MAY-2000.

XX

PF 15-SEP-1997; 97US-0937271.

XX

PR 16-SEP-1992; 92US-0945954.

XX

PA (UYTE-) UNIV TENNESSEE RES CORP.

PI Lederer JW, Dale JB;

XX

DR WPI; 2000-364475/31.

DR N-PSDB; A57893.

XX

XX New immunogenic recombinant hybrid M protein comprising amino-terminal
peptide fragments of streptococcal M protein useful as vaccine against
rheumatic fever and infections leading to rheumatic fever -

XX

PS Disclosure; Fig 1A-B; 62pp; English.

XX

CC The invention relates to multivalent immunogenic hybrid group A
streptococcal M proteins comprising N-terminal peptide fragments of M
proteins that elicit opsonic antibodies against multiple serotypes
of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
generated using the hybrid proteins are against one or more M protein
serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
The invention also encompasses a recombinant DNA molecule comprising a
nucleotide sequence that encodes a multivalent hybrid M protein; and a
method for immunising a mammal against streptococcal infections,
comprising administering an immunogenic multivalent hybrid M protein to
the mammal. The multivalent hybrid M proteins are useful for eliciting
opsonic or protective antibodies to the M proteins of different serotypes
of group A streptococci, and may therefore be used as vaccines to protect
against and control infection by type A streptococci. Type A streptococci
are not only responsible for streptococcal pharyngitis (strep throat),
forms of pneumonia and a condition resembling toxic shock, but are also
involved in the development of acute rheumatic fever (ARF) and rheumatic
heart disease. In a patient with ARF, antibodies formed during a group A
streptococcal infection are also cross-reactive with heart tissue, which
indicates that the streptococci and host tissue contain similar antigenic
motifs. The new multivalent vaccines are capable of raising sero-specific
antibodies against various serotypes of group A streptococci which are
not cross-reactive with human heart tissue. Sequences B03113-B03117,
B03119-B03121 and B03123-B03124 represent multivalent hybrid
Streptococcus pyogenes M proteins generated in the disclosure of the
invention.

XX Sequence 287 AA;

QY 304 VATRSOTDTLEKVQRADKFEIENNTKLKNSDLSFNKALKDHNDLTEELSNAKEKL 363
|||||

Db 2 vatrsqtdtlekvqradkfeienntklknsdlsfnkalkdhndelteelsnakeklr 61

QY 364 KNDKSLSEKASKIQELEAR 383
|||||

Db 62 kndkselsekaskiqelear 81

RESULT 6

R50998

ID R50998 standard; Protein; 173 AA.

XX

AC R50998;

XX

DT 02-NOV-1994 (first entry)

XX

DE Recombinant M24-M5 divalent hybrid.

XX

XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
M protein; M24; M5; M6; M19; subunit; tetravalent protein; emm gene;
tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
restriction enzyme site; multivalent M protein; immunisation; group A;
streptococci; rheumatic fever; rheumatic heart disease; humoral;
antibody; heart tissue; antigen; serotype; mucosal.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 2..115

FT Peptide /label= M24

FT Peptide 116..173

FT Peptide /label= M5

XX

PN W09406421-A.

XX

PD 31-MAR-1994.

XX

PF 15-SEP-1993; 93WO-US08703.

XX

PR 16-SEP-1992; 92US-0945954.

XX

PA (UYTE-) UNIV TENNESSEE RES CORP.

XX

PI Dale JB, Lederer JW;

XX

DR WPI; 1994-118122/14.

DR N-PSDB; Q45220.

XX

XX New immunogenic hybrid proteins derives from streptococcal M proteins
- induces opsonic antibodies, for protective immunisation against
against multiple group A streptococci serotypes

PS Disclosure; Fig 10; 67pp; English.

XX

CC The sequences given in R50992-1001 represent hybrid M proteins which
contain the M24-M5-M6 and/or M19 subunits. These multivalent
proteins were constructed using fragments of the 5' regions of emm
genes that were amplified by PCR, ligated in tandem and expressed in
CC tissue-cross-reactive epitopes, which can then be linked into one
protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into

CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.

XX
 SQ Sequence 173 AA;

Query Match 18.8%; Score 72; DB 15; Length 173;
 Best Local Similarity 100.0%; Pred. No. 2.4e-57;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVRADKFEIENNTLKNKSDLSFNKALKDHNDLTELSSNAKEKLR 363
 |||||
 Db 2 vatrsqdtlekvrqradkfeienntlkksdlsfnnkalkdhndelteelsnakeklr 61

Qy 364 KNDKSLSEKASK 375
 |||||
 Db 62 kndkslseask 73

RESULT 7
 B03120
 ID B03120 standard; Protein; 173 AA.

XX AC B03120;

XX DT 10-OCT-2000 (first entry)

XX DE S. pyogenes hybrid M protein (M24-M5), SEQ ID NO:15.

XX KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.

XX OS Streptococcus pyogenes.

OS Synthetic.

XX PN US6063386-A.

XX PD 16-MAY-2000.

XX PF 15-SEP-1997; 97US-0937271.

XX PR 16-SEP-1992; 92US-0945954.

XX PA (UYTE-) UNIV TENNESSEE RES CORP.

PI Lederer JW, Dale JB;

XX WPI; 2000-364475/31.

DR N-PSDB; A57899.

XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever

XX PS Disclosure: Fig 10A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
 CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
 CC Streptococcus pyogenes M proteins generated in the disclosure of the
 CC invention.

XX
 SQ Sequence 173 AA;

Query Match 18.8%; Score 72; DB 21; Length 173;
 Best Local Similarity 100.0%; Pred. No. 2.4e-57;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVRADKFEIENNTLKNKSDLSFNKALKDHNDLTELSSNAKEKLR 363
 |||||
 Db 2 vatrsqdtlekvrqradkfeienntlkksdlsfnnkalkdhndelteelsnakeklr 61

Qy 364 KNDKSLSEKASK 375
 |||||
 Db 62 kndkslseask 73

RESULT 8

R50997

ID R50997 standard; Protein; 247 AA.

XX AC R50997;

XX DT 02-NOV-1994 (first entry)

XX DE Recombinant M19-M6-M4-M24 tetraivalent hybrid.

XX KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; pKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal.

XX OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 2..38

FT Peptide /label= M19

FT Peptide 39..75

FT Peptide /label= M6

FT Peptide 76..135

FT Peptide /label= M5

FT Peptide 136..247

FT Peptide /label= M24

XX WO9406421-A.

XX 31-MAR-1994.

XX 15-SEP-1993; 93WO-US08703.

XX 16-SEP-1992; 92US-0945954.

XX (UYTE-) UNIV TENNESSEE RES CORP.

Tue May 15 07:28:29 2001

us-09-151-409-16.oligo.rag

```

XX PI Dale JB, Lederer JW;
XX WPI; 1994-118122/14.
DR N-PSDB; Q45219.
XX
XX New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes
XX
XX Disclosure; Fig 9; 67pp; English.
XX
XX The sequences given in R50992-1001 represent hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.
XX
XX Sequence. 247 AA;
SQ
Query Match 18.8%; Score 72; DB 15; Length 247;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 304 VATRSQTDLEKVRADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 136 vatrsqtdtlekvradkfeienttlkksdlsfnkalkdhndlteelsnakeklr 195
QY 364 KNDKSLSEKASK 375
DB 196 kndkslsekask 207
RESULT 9
B03119 ID B03119 standard; Protein; 247 AA.
XX AC B03119;
XX 10-OCT-2000 (first entry)
XX S. pyogenes hybrid M protein (M19-M6-M5-M24), SEQ ID NO:13.
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia.
XX Streptococcus pyogenes.
OS Synthetic.
XX US6063386-A.
XX 16-MAY-2000.
XX 15-SEP-1997; 97US-0937271.
XX 16-SEP-1992; 92US-0945954.
PR

```

```

XX PA (UYTE-) UNIV TENNESSEE RES CORP.
XX PI Lederer JW, Dale JB;
XX WPI; 2000-364475/31.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX
XX Disclosure; Fig 9A-B; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
CC invention.
XX
XX Sequence 247 AA;
SQ
Query Match 18.8%; Score 72; DB 21; Length 247;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 304 VATRSQTDLEKVRADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 136 vatrsqtdtlekvradkfeienttlkksdlsfnkalkdhndlteelsnakeklr 195
QY 364 KNDKSLSEKASK 375
DB 196 kndkslsekask 207
RESULT 10
R50994 ID R50994 standard; Protein; 274 AA.
XX AC R50994;
XX 02-NOV-1994 (first entry)
XX Recombinant M24-M5-M6-M19 (linker variant).
XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal.
XX

```


OS Synthetic.

XX Key Location/Qualifiers
 FT Peptide 1..115
 FT Peptide /label= M24
 FT Peptide 116..124
 FT Peptide /label= Linker
 FT Peptide 125..184
 FT Peptide /label= M5
 FT Peptide 185..193
 FT Peptide /label= Linker
 FT Peptide 194..230
 FT Peptide /label= M6
 FT Peptide 231..239
 FT Peptide /label= Linker
 FT Peptide 240..274
 FT Peptide /label= M5

XX WO9406421-A.
 PN 31-MAR-1994.
 XX 15-SEP-1993; 93WO-US08703.
 XX 16-SEP-1992; 92US-0945954.
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 PA Dale JB, Lederer JW;
 PI WPI; 1994-118122/14.
 DR N-PSDB; 045216.

XX New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX Disclosure; Fig 6; 67pp; English.

XX The sequences given in R50992-1001 represent hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC PKK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.

XX Sequence 274 AA;

Query Match 18.8%; Score 72; DB 15; Length 274;
 Best Local Similarity 100.0%; Pred. No. 3.4e-57;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVOERADFEFENNTLKNKSDLSFNKALKDHPNDELTEELSNKELR 363
 |||||||
 Db 2 vatrsqdtlekvqrdrkfennltklknsdlsfnkalkdhdteelsnakeklr 61

QY 364 KNDXSLSEKASK 375

Db 62 knkdsalsekask 73
 |||||||

RESULT 11

B03115
 ID B03115 standard; Protein; 274 AA.

XX AC B03115;

XX DT 10-OCT-2000 (first entry)

XX DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:6.

XX Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 XX pneumonia.

OS Streptococcus pyogenes.

OS Synthetic.

XX PN US6063386-A.

XX PD 16-MAY-2000.

XX PF 15-SEP-1997; 97US-0937271.

XX PR 16-SEP-1992; 92US-0945954.

XX PA (UYTE-) UNIV TENNESSEE RES CORP.

XX PI Lederer JW, Dale JB;

XX DR WPI; 2000-364475/31.

XX N-PSDB; A57895.

XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever

XX Disclosure; Fig 6A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infection,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
 CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
 CC Streptococcus pyogenes M proteins generated in the disclosure of the
 CC invention.

XX Sequence 274 AA;

Query Match 18.8%; Score 72; DB 21; Length 274;
 Best Local Similarity 100.0%; Pred. No. 3.4e-57;

Matches	72: Conservative	0: Mismatches	0: Indels	0: Gaps	0:
QY	304	VATRSQTDLTLEKVGQRADKFEIENNVLKLNKSDLSFNKKALKDHNDELTEELSSNAKEKLR	363		
DB	2	vatrsgtdtlekvgqeradkfeientiklknssdlfnkkalkdhnldtelteelsnakeklr	61		
QY	364	KNDKSLSEKASK	375		
DB	62	kndkslsekask	73		
RESULT	12				
R50999					
ID	R50999	standard; Protein; 343 AA.			
XX	R50999;				
XX	02-NOV-1994	(first entry)			
XX	Recombinant tetraivalent-C repeat.				
XX	Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;				
KW	M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;				
KW	tandem; pKK223.3; protective epitope; tissue-cross-reactive epitope;				
KW	restriction enzyme site; multivalent M protein; immunisation; group A;				
KW	streptococci; rheumatic fever; rheumatic heart disease; humoral;				
KW	antibody; heart tissue; antigen; serotype; mucosal.				
XX	Synthetic.				
XX	Key	Location/Qualifiers			
FH	Peptide	1..38			
FT	Peptide	/label= M19			
FT	Peptide	39..75			
FT	Peptide	/label= M6			
FT	Peptide	76..135			
FT	Peptide	/label= M5			
FT	Peptide	136..249			
FT	Peptide	/label= M24			
FT	Peptide	250..343			
FT	Peptide	/label= C-terminal_repeats			
XX	W09406421-A.				
XX	31-MAR-1994.				
XX	15-SEP-1993;	93WO-US08703.			
XX	16-SEP-1992;	92US-0945954.			
XX	(UYTE-) UNIV TENNESSEE RES CORP.				
XX	Dale JB, Lederer JW;				
XX	WPI; 1994-118122/14.				
XX	N-PSDB; Q45221.				
XX	New immunogenic hybrid proteins derives from streptococcal M proteins				
XX	induces opsonic antibodies, for protective immunisation against				
XX	against multiple group A streptococci serotypes				
XX	Disclosure; Fig 11; 67pp; English.				
XX	The sequences given in R50992-1001 represent hybrid M proteins which				
XX	contain the M24-M5-M6 and/or M19 subunits. These multivalent				
XX	proteins were constructed using fragments of the 5' regions of emm				
XX	genes that were amplified by PCR, ligated in tandem and expressed in				
XX	pKK223.3. The amplified regions pref. encode protective and not				
XX	tissue-cross-reactive epitopes, which can then be linked into one				
XX	protein molecule. The recombinant hybrid protein may contain 113				
XX	N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and				
XX	35 from M19. Each section is linked by 2 amino acids specified by				
XX	the respective restriction enzyme sites that were synthesised into				

method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences B03113-B03117, B03119-B03121 and B03123-B03124 represent multivalent hybrid Streptococcus pyogenes M proteins generated in the disclosure of the invention.

SQ Sequence 343 AA;

Query Match 18.8%; Score 72; DB 21; Length 343;

Best Local Similarity 100.0%; Pred. No. 4.le-57; Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDELTELSNAKEKLR 363

Db 136 vatrsqdtlekvgeradkfeienntklknsdlsfnkalkdhndeiteelsnakeklr 195

Qy 364 KNDKSLSEKASK 375

Db 196 kndkslsekask 207

RESULT 14

ID P50295 standard; Protein; 100 AA.

XX P50295;

XX P50295;

DT 01-DEC-1991 (first entry)

XX Sequence encoded by a portion of the emm6 gene that encodes the amino terminal end of the M6 protein.

XX Fibrillar surface molecule; rheumatic fever; glomerulonephritis; throat swabs; diagnosis.

XX Streptococcus pyogenes (Group A streptococcus) strain D471.

XX Key Location/Qualifiers

FT Protein 43..100

FT /label= N-terminal of M6

XX W08500832-A.

XX 28-FEB-1985.

XX 09-AUG-1984; 84WO-4001261.

XX 18-JUN-1984; 84US-0521716.

XX 10-AUG-1983; 83US-0521962.

XX (UYRO-) ROCKEFELLER UNIV.

XX (UYEM-) EMORY UNIV.

XX (UYEM-) EMORY UNIV.

XX Scott JR, Fischetti VA;

XX WPI; 1985-062291/10.

XX N-PSDB; N50341.

XX

PT Immunogenic streptococcal polypeptide prodn. - by recombinant DNA methods, useful as vaccines and as diagnostic probes

XX Example: Fig 4; 46pp; English.

XX The inventors claim a polypeptide having an immunoreactive and antigenic determinant of a Streptococcus pyogenes M protein produced by a unicellular organism, and a DNA sequence encoding it. Also claimed are purified DNA probes able to bind to the S.pyogenes gene which codes for M protein. The polypeptides can be used in vaccines.

XX Sequence 100 AA;

Query Match 9.1%; Score 35; DB 6; Length 100;

Best Local Similarity 100.0%; Pred. No. 3.6e-24;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 RVFPRGTVENPKARELLNKYDVENSMLQANNNDKL 160

Db 43 rvfprgtvenpkarelknkydvensmlqannndkl 77

RESULT 15

P90955

ID P90955 standard; protein; 441 AA.

XX AC P90955;

XX AC P90955;

DT 23-FEB-1990 (first entry)

XX M6 streptococcal protein.

XX Immunoglobulin.

XX Streptococcus group A.

XX Key Location/Qualifiers

FT Region 216..235

FT Region 248..269

FT Region 275..284

XX W08909064-A.

XX 05-OCT-1989.

XX 13-MAR-1989; 89WO-US01026.

XX 25-MAR-1988; 88US-0173380.

XX (UYRO) ROCKEFELLER UNIV.

XX Fischetti V;

XX WPI; 1989-309382/42.

XX Polypeptide(s) from streptococcal M protein - used to prepare vaccines for providing protection against streptococcal infection

XX Disclosure: Fig 1; 22pp; English.

XX The regions (pref. conjugated to a natural carrier, eg cholera toxin) above can elicit an secretory Ig response in a mammal. They are used in vaccines against streptococcal infection, and give protection to different serotypes.

XX Sequence 441 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 1.2e-23;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 RVFPRGTVENPDKARELLNKYDVENSMLQANNDKL 160
|||||
Db 1 rvfprgvtvenpdkarellnkydvensmlqandkl 35

Search completed: May 10, 2001, 08:40:47
Job time: 149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:38:53 ; Search time 13.07 Seconds
(without alignments)
562.950 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKSKIQLEARK 383

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 185757 seqs, 19210857 residues

Word size : 10

Total number of hits satisfying chosen parameters: 11

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	134	35.0		247	3	US-08-937-271-4	Sequence 4, Appl
2	80	20.9		287	3	US-08-937-271-2	Sequence 2, Appl
3	72	18.8		173	3	US-08-937-271-15	Sequence 15, Appl
4	72	18.8		247	3	US-08-937-271-13	Sequence 13, Appl
5	72	18.8		274	3	US-08-937-271-6	Sequence 6, Appl
6	72	18.8		343	3	US-08-937-271-17	Sequence 17, Appl
7	50	13.1		443	2	US-08-795-475-6	Sequence 6, Appl
8	18	4.7		67	3	US-08-937-271-20	Sequence 20, Appl
9	18	4.7		135	3	US-08-937-271-22	Sequence 22, Appl
10	18	4.7		187	3	US-08-937-271-8	Sequence 8, Appl
11	18	4.7		305	3	US-08-937-271-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-937-271-4
; Sequence 4, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/937,271
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/937,271
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:

Query Match 35.0%; Score 134; DB 3; Length 247;
Best Local Similarity 100.0%; Pred. No. 8.6e-123;
Matches 134; Conservative 0; Mismatches 0; Gaps 0;

QY 64 GSAVTRGTINDPQRAKALDKYLENHDLKTKNEGLKTKNEGLKTKNEGLKTE 123
Db 114 GSAVTRGTINDPQRAKALDKYLENHDLKTKNEGLKTKNEGLKTKNEGLKTE 173
QY 124 VDRVFRGTVENPDKARELKNKYDVNSMLQANNDKLPWRVRYTRHTPDK 183
Db 174 VDRVFRGTVENPDKARELKNKYDVNSMLQANNDKLPWRVRYTRHTPDK 233
QY 184 AKEHELQOQNEKLS 197
Db 234 AKEHELQOQNEKLS 247

RESULT 2
US-08-937-271-2
; Sequence 2, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/937,271
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-2

Query Match 20.9%; Score 80; DB 3; Length 287;
Best Local Similarity 100.0%; Pred. No. 3.6e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 2 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 61

QY 364 KNDKSLSEKASKIQLK 383
DB 62 KNDKSLSEKASKIQLK 81

RESULT 3
US-08-937-271-15
; Sequence 15, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-15

Query Match 18.8%; Score 72; DB 3; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.9e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 2 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 61

QY 364 KNDKSLSEKASKIQLK 383
DB 62 KNDKSLSEKASKIQLK 81

RESULT 5
US-08-937-271-6
; Sequence 6, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
```

```

Best Local Similarity 100.0%; Pred. No. 1.4e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 2 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 61

QY 364 KNDKSLSEKASKIQLK 375
DB 62 KNDKSLSEKASKIQLK 73

RESULT 4
US-08-937-271-13
; Sequence 13, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-13

Query Match 18.8%; Score 72; DB 3; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 136 VATRSQDTLEKVVQERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 195

QY 364 KNDKSLSEKASKIQLK 375
DB 196 KNDKSLSEKASKIQLK 207

RESULT 5
US-08-937-271-6
; Sequence 6, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
```

; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-6

Query Match 18.8%; Score 72; DB 3; Length 274;
Best Local Similarity 100.0%; Pred. No. 2.1e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 363
|||||
DB 2 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 61
|||||
QY 364 KNDKSLSEKASK 375
|||||
DB 62 KNDKSLSEKASK 73
|||||

RESULT 6
US-08-937-271-17
; Sequence 17, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271

; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-17

Query Match 18.8%; Score 72; DB 3; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.6e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 363
|||||
DB 136 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 195
|||||
QY 364 KNDKSLSEKASK 375
|||||
DB 196 KNDKSLSEKASK 207
|||||

RESULT 7
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: Sjobring, Ulf
; TITLE OF INVENTION: PROTEIN L-AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McWaters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-475-6

Query Match 13.1%; Score 50; DB 2; Length 443;

```
Best Local Similarity 100.0%; Pred. No. 8.4e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPAIQNIRLHENKDLKARLENAMVAGRDFKRA 249
Db 1 NGDGNPREVIEDLAANNPAIQNIRLHENKDLKARLENAMVAGRDFKRA 50

RESULT 8
US-08-937-271-20
; Sequence 20, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 67 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-20

Query Match 4.7%; Score 18; DB 3; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 EVDVFPFGTGVNPKAR 140
Db 33 EVDVFPFGTGVNPKAR 50

RESULT 9
US-08-937-271-22
; Sequence 22, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
```

```
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-22

Query Match 4.7%; Score 18; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 EVDVFPFGTGVNPKAR 140
Db 33 EVDVFPFGTGVNPKAR 50

RESULT 10
US-08-937-271-8
; Sequence 8, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 187 amino acids
```


; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-8

Query Match 4.7%; Score 18; DB 3; Length 187;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 64 GSAVTRGTINDPORAKEA 81
Db 47 GSAVTRGTINDPORAKEA 64

RESULT 11
US-08-937-271-10
; Sequence 10, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-10

Query Match 4.7%; Score 18; DB 3; Length 305;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 123 EVDVFPFGTGVNPKAR 140
Db 33 EVDVFPFGTGVNPKAR 50

Search completed: May 10, 2001, 08:41:04
Job time: 131 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:39:43 ; Search time 18.37 Seconds
(without alignments)
1432.819 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEAR 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 198801 seqs, 68722935 residues

Word size : 10

Total number of hits satisfying chosen parameters: 14

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR_67: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	20.9	539	2	A28549 M24 protein precurs
2	58	15.1	492	2	A28616 M5 protein precurs
3	50	13.1	197	1	MMSOMP M5 protein - Strep
4	50	13.1	227	2	S01260 M1 protein precurs
5	50	13.1	484	2	S35401 M1 protein precurs
6	50	13.1	484	2	S46489 M protein precurs
7	36	9.4	83	2	S61086 M protein precurs
8	35	9.1	98	2	S61079 M protein precurs
9	35	9.1	150	2	S60838 M protein - Strep
10	35	9.1	483	2	A26297 M protein type 19
11	24	6.3	96	2	S71519 M protein precurs
12	18	4.7	87	2	S60786 M protein pepM19
13	16	4.2	26	2	PL0027 M protein precurs
14	14	3.7	87	2	S61075 M protein precurs

ALIGNMENTS

RESULT 1
A28549 M24 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M24
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 26-Aug-1999
C:Accession: A28549; S60802
R:Howe, A.R.; Beachey, E.H.; Burdett, V.
J. Bacteriol. 170, 676-684, 1988
A:Title: Molecular evolution of streptococcal M protein: cloning and nucleotide sequence

A:Reference number: A28549; MUID:88115166
A:Accession: A28549
A:Molecule type: DNA
A:Residues: 1-539 <MO>
R:Cross-references: GB:M19031; NID:g153616; PIDN:AAA26874.1; PID:g153617
R:Whatmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60802
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 30-89 <WHA>
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein

Query Match 20.9%; Score 80; DB 2; Length 539;
Best Local Similarity 100.0%; Pred. No. 6.8e-72;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALDHNDLFEELSNAKEKLR 363
|||||
Db 43 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALDHNDLFEELSNAKEKLR 102
|||||

Qy 364 KNDKSLSEKASKIQELEAR 383
|||||
Db 103 KNDKSLSEKASKIQELEAR 122
|||||

RESULT 2

A28616 M5 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M5
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 26-Aug-1999
C:Accession: A28616; S60787
R:Miller, L.; Gray, L.; Beachey, E.; Kehoe, M.
J. Biol. Chem. 263, 5668-5673, 1988
A:Title: Antigenic variation among group A streptococcal M proteins. Nucleotide sequence
A:Reference number: A28616; MUID:88186881
A:Accession: A28616
A:Molecule type: DNA
A:Residues: 1-492 <ML>
A:Cross-references: GB:M20374; NID:g153812; PIDN:AAA26976.1; PID:g153813
R:Whatmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60787
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 30-89 <WHA>
C:Genetics:
A:Gene: smp5
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein
F:1-42/Domain: signal sequence #status predicted <SIG>
F:43-492/Product: M5 protein #status predicted <MAT>

Query Match 15.1%; Score 58; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 7.1e-50;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 AVTRGINDPQAKALKDYELNHDLTKNKEGLKTENEGKLTENEGKLTKE 123
|||||
Db 43 AVTRGINDPQAKALKDYELNHDLTKNKEGLKTENEGKLTENEGKLTKE 100
|||||

RESULT 3

MMSOMP M5 protein - Streptococcus pyogenes (fragment)

```

C:Species: Streptococcus pyogenes
C:Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Feb-1997
C:Accession: A03501
R:Manjula, B.N.; Acharya, A.S.; Mische, S.M.; Fairwell, T.; Fischetti, V.A.
J. Biol. Chem. 259, 3686-3693, 1984
A:Title: The complete amino acid sequence of a biologically active 197-residue fragment
A:Reference number: A03501; MUID:84162039
A:Accession: A03501
A:Molecule type: protein
A:Residues: 1-197 <MAN>
C:Comment: Residues 27-58 contain several tandem 7-residue repeats, and residues 150-172
C:Comment: This protein is closely associated with the virulence of the bacterium and it
C:Superfamily: M5 protein
C:Keywords: Cell wall; transmembrane protein; virulence

Query Match 13.1%; Score 50; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.1e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 DPQRAEALDKYLENHDLKTKNEGLKTENEGLKTENEGLKTE 123
Db 9 DPQRAEALDKYLENHDLKTKNEGLKTENEGLKTENEGLKTE 58

```

RESULT 4
 M protein precursor - Streptococcus pyogenes (serotype M1) (fragment)
 S01260
 C:Species: Streptococcus pyogenes
 C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
 C:Accession: S01260; S00767
 R:Hannes-Fritz, E.
 submitted to the EMBL data Library, June 1988
 A:Reference number: S01260
 A:Accession: S01260
 A:Molecule type: DNA
 A:Residues: 1-227 <HAN>
 A:Cross-references: EMBL:X07933; NID:g47358; PIDN:CAA30757.1; PID:g47359
 A:Experimental source: strain CS130
 R:Haanes-Fritz, E.; Kraus, W.; Burdett, V.; Dale, J.B.; Beachey, E.H.; Cleary, P.
 Nucleic Acids Res. 16, 4667-4677, 1988
 A:Title: Comparison of the leader sequences of four group A streptococcal M protein genes
 A:Reference number: S00767; MUID:88247768
 A:Accession: S00767
 A:Molecule type: DNA
 A:Residues: 1-199 <HAA>
 A:Cross-references: EMBL:X07860
 A:Experimental source: strain CS130
 C:Genetics:
 A:Gene: emm1
 C:Superfamily: M5 protein
 F:1-41/Domain: signal sequence #status predicted <SIG>
 F:42-227/Product: M protein #status predicted <MAT>

```

Query Match          13.1%; Score 50; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 3.6e-42;      0;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    200  NGDGNPREVIEDLAANNPAIQNIRLRHNKDKLKA RL ENAMEVAGRDPFKRA 249
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB     42  NGDGNPREVIEDLAANNPAIQNIRLRHNKDKLKA RL ENAMEVAGRDPFKRA 91

RESULT      5
S35401
M1 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M1
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S35401; S61074; S60784
R:Pdbielecki, A.
submitted to the EMBL Data Library, September 1991

```

```

RESULT 7
S61086
M protein precursor - Streptococcus pyogenes (serotype M68) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M68
C:Date: 15-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 16-Feb-1997
C:Accession: S61086; S60825
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
A:Description: Noncongruent relationships between variation in emml gene sequences and the
A:Reference number: S61072
A:Accession: S61086
A:Molecule type: DNA
A:Residues: 1-83 <WHA>
A:Cross-references: EMBL:U11997
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the po
A:Reference number: S60784; MUID:95198537
A:Accession: S60825
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 18-43,'A',45-78 <WHW>
A:Cross-references: EMBL:U11997
C:Genetics:
A:Gene: emm68
C:Superfamily: M5 protein
F:1-30/Domain: signal sequence (fragment) #status predicted <SIG>
F:31-83/Product: M protein (fragment) #status predicted <MAT>

Query Match          9.4%; Score 36; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 ANPAQIQLRLHENKDLKARLENAMVAGRDFKRA 249
      |||||
Db 45 ANPAQIQLRLHENKDLKARLENAMVAGRDFKRA 80

RESULT 8
S61079
M protein precursor - Streptococcus pyogenes (serotype M19) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M19
C:Date: 15-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 16-Feb-1997
C:Accession: S61079; S60799
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
A:Description: Noncongruent relationships between variation in emml gene sequences and the
A:Reference number: S61072
A:Accession: S61079
A:Molecule type: DNA
A:Residues: 1-98 <WHA>
A:Cross-references: EMBL:U11959
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the po
A:Reference number: S60784; MUID:95198537
A:Accession: S60799
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 18-77 <WHW>
A:Cross-references: EMBL:U11959
C:Genetics:
A:Gene: emm19
C:Superfamily: M5 protein
F:1-30/Domain: signal sequence (fragment) #status predicted <SIG>
F:31-98/Product: M protein (fragment) #status predicted <MAT>

Query Match          9.1%; Score 35; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.8e-27;

```

```

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRYTRHTPEDKLIKIIDDLDAKEHELOQQNEKLS 197
      |||||
Db 31 RVRYTRHTPEDKLIKIIDDLDAKEHELOQQNEKLS 65

RESULT 9
S60838
M protein precursor - Streptococcus pyogenes (serotype M6) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M6
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60838
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60838
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-150 <WHA>
A:Cross-references: EMBL:U11986; NID:G533649; PIDN:AAA99602.1; PID:G1235835
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Superfamily: M5 protein

Query Match          9.1%; Score 35; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.6e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
      |||||
Db 30 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 64

RESULT 10
A26297
M6 protein - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 26-Aug-1999
C:Accession: A26297
R:Hollingshead, S.K.; Fischetti, V.A.; Scott, J.R.
J. Biol. Chem. 261, 1677-1686, 1986
A:Title: Complete nucleotide sequence of type 6 M protein of the group A streptococcu
A:Reference number: A26297; MUID:86111835
A:Accession: A26297
A:Molecule type: DNA
A:Residues: 1-483 <HOL>
A:Cross-references: GB:M11338; GB:M11415; NID:G153699; PIDN:AAA26920.1; PID:G153700
C:Genetics:
A:Gene: emm6
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein

Query Match          9.1%; Score 35; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 7.9e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
      |||||
Db 43 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 77

RESULT 11
S71519
M protein type 19 - Streptococcus pyogenes (fragment)
C:Species: Streptococcus pyogenes
C:Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
C:Accession: S71519; S71515
R:Podbielski, A.; Meizer, B.; Luttkien, R.
submitted to the EMBL Data Library, November 1990

```

A:Description: Application of the polymerase chain reaction to study the M protein(-like)
A:Reference number: S71518
A:Accession: S71519
A:Molecule type: DNA
A:Residues: 1-96 <POD>
A:Cross-references: EMBL:X56609
A:Experimental source: strain J17D/70
R:Podbielski, A.; Melzer, B.; Lueticken, R.
Med. Microbiol. Immunol. 180, 213-227, 1991
A:Title: Application of the polymerase chain reaction to study the M protein(-like) gene
A:Reference number: S71514; MUID:92149493
A:Accession: S71515
A:Molecule type: DNA
A:Residues: 1-42 <POW>
A:Cross-references: EMBL:X56609
A:Experimental source: J17D/70
C:Genetics:
A:Gene: emm19
C:Superfamily: M5 protein

Query Match 6.3%; Score 24; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRVTRHTPEDKLLKIIDDLDAKE 186
|||||
Db 43 RVRVTRHTPEDKLLKIIDDLDAKE 66

RESULT 12
M protein precursor - Streptococcus pyogenes (serotype M3) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M3
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S60786
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the pop
A:Reference number: S60784; MUID:95198537
A:Accession: S60786
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-87 <WHA>
A:Cross-references: EMBL:U11945; NID:g533567; PIDN:AAA99561.1; PID:g533568
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994

Query Match 4.7%; Score 18; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 AQAGRLDLRQAEYLYKGL 289
|||||
Db 70 AQAGRLDLRQAEYLYKGL 87

RESULT 13
M protein pepM19 - Streptococcus sp. (fragment)
C:Species: Streptococcus sp.
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Feb-1997
C:Accession: PL0027
R:Bronze, M.S.; Beachey, E.H.; Dale, J.B.
J. Exp. Med. 167, 1849-1859, 1988
A:Title: Protective and heart-crossreactive epitopes located within the NH2 terminus of
A:Reference number: PL0027; MUID:88258373
A:Accession: PL0027
A:Molecule type: protein
A:Residues: 1-26 <BRO>
C:Superfamily: M5 protein
C:Keywords: virulence factor

Query Match 4.2%; Score 16; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 5.6e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 PEDKLLKIIDDLDAKE 186
|||||
Db 8 PEDKLLKIIDDLDAKE 23

RESULT 14
M protein precursor - Streptococcus pyogenes (serotype M31) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M31
C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 15-Oct-1999
C:Accession: S61075; S60809
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
submitted to the EMBL Data Library, July 1994
A:Description: Noncongruent relationships between variation in emm gene sequences an
A:Reference number: S61072
A:Accession: S61075
A:Molecule type: DNA
A:Residues: 1-87 <WHA>
A:Cross-references: EMBL:U11943; NID:g533563; PIDN:AAA99559.1; PID:g1235811
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60809
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 9-64, 'P' <WHW>
A:Cross-references: EMBL:U11943
C:Genetics:
A:Gene: emm31
F;1-16/Domain: signal sequence (fragment) #status predicted <SIG>
F;17-87/Product: M protein (fragment) #status predicted <MAT>

Query Match 3.7%; Score 14; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 QKAEYLYKGLNDWAE 294
|||||
Db 69 QKAEYLYKGLNDWAE 82

Search completed: May 10, 2001, 08:41:26
Job time: 103 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:41:08 ; Search time 11.87 Seconds
(without alignments)
1105.295 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVKQERA.....KNDKSLSEKASKIQELEAR 383

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 10

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	80	20.9	539	1 M24_STRPY	P12379 streptococc
2	58	15.1	492	1 M5_STRPY	P02977 streptococc
3	35	9.1	483	1 M6_STRPY	P08089 streptococc

ALIGNMENTS

RESULT 1

M24_STRPY	1	80	20.9	539	1 M24_STRPY	P12379 streptococc
ID	M24_STRPY	STANDARD;	PRT;	539 AA.		
AC	P12379;					
DT	01-OCT-1989 (Rel. 12, Created)					
DT	01-OCT-1989 (Rel. 12, Last sequence update)					
DT	01-OCT-1996 (Rel. 34, Last annotation update)					
DE	M PROTEIN, SEROTYPE 24 PRECURSOR.					
GN	EMM24.					
OS	Streptococcus pyogenes.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;					
CC	Streptococcus.					
OX	NCBI_TaxID=1314;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=A24 VAUGHN;					
RX	MEDLINE=88115166; PubMed=3276665;					
RA	Mouw A.R., Beachey E.H., Burdett V.;					
RT	"Molecular evolution of streptococcal M protein: cloning and					
RT	nucleotide sequence of the type 24 M protein gene and relation to					
RT	other genes of Streptococcus pyogenes.";					
RL	J. Bacteriol. 170:676-684(1988).					
CC	-1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES					
CC	OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF					
CC	THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO					

PHAGOCYTOSIS.

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.

-1- SIMILARITY: TO OTHER M PROTEINS.

-1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS IN THE REGION OF THE MEMBRANE ANCHOR.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M19031; AAA26874.1; -

PIR; A28549; A28549.

InterPro; IPR001899; -

Pfam; PF00746; Gram_pos_anchor; 1.

PRINTS; PR00015; GPOSANCHOR.

PROSITE; PS00343; GRAM_POS_ANCHORING; 1.

Signal; Virulence; Cell wall; Transmembrane; Phagocytosis; Duplication; Repeat; Antigen; Coiled coil.

SIGNAL 1 42

FT CHAIN 43 539 M PROTEIN, SEROTYPE 24.

FT DOMAIN 43 514 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 515 534 MEMBRANE ANCHOR.

FT DOMAIN 535 539 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 118 301 5.3 X 35 AA TANDEM REPEATS, A-TYPE.

FT REPEAT 118 152 A-1.

FT REPEAT 153 187 A-2 (1 ALTERATION).

FT REPEAT 188 222 A-3 (2 ALTERATIONS).

FT REPEAT 223 257 A-4 (1 ALTERATION).

FT REPEAT 258 292 A-5 (2 ALTERATIONS).

FT REPEAT 293 301 A-6 (INCOMPLETE).

FT DOMAIN 311 405 2.7 X 35 AA TANDEM REPEATS, B-TYPE.

FT REPEAT 311 355 B-1.

FT REPEAT 356 380 B-2.

FT REPEAT 381 405 B-3 (INCOMPLETE).

FT DOMAIN 468 504 GLY/PRO-RICH (CELL WALL-SPANNING).

FT DOMAIN 505 510 CONSERVED IN GRAM-POSITIVE COCCI SURFACE PROTEINS.

SQ SEQUENCE 539 AA; 58804 MW; B03EDF3AC1E6E9C7 CRC64;

Query Match 20.9%; Score 80; DB 1; Length 539;

Best Local Similarity 100.0%; Pred. No. 6.2e-70;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVKQERADKFEIENNTLKLNSDLSFNKKALDHNDLTELSENAKEKL 363

DB 43 VATRSQDTLEKVKQERADKFEIENNTLKLNSDLSFNKKALDHNDLTELSENAKEKL 102

QY 364 KNDKSLSEKASKIQELEAR 383

DB 103 KNDKSLSEKASKIQELEAR 122

RESULT 2

M5_STRPY

ID	M5_STRPY	STANDARD;	PRT;	492 AA.
AC	P02977;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	M PROTEIN, SEROTYPE 5 PRECURSOR.			
GN	EMM5 OR SMP5.			
OS	Streptococcus pyogenes.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
CC	Streptococcus.			
OX	NCBI_TaxID=1314;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88186881; PubMed=3281944;			

RA Miller L., Gray L., Beachey E., Kehoe M.;
RT "Antigenic variation among group A streptococcal M proteins.
RT Nucleotide sequence of the serotype 5 M protein gene and its
RT relationship with genes encoding types 6 and 24 M proteins.";
RL J. Biol. Chem. 263:5668-5673(1988).
RN [2]
RN SEQUENCE OF 43-212 AND 238-250.
RX MEDLINE-84162039; PubMed-6368549;
RA Manjula B.N., Acharya A.S., Mische S.M., Fairwell T., Fischetti V.A.;
RT "The complete amino acid sequence of a biologically active
RT 197-residue fragment of M protein isolated from type 5 group A
RT streptococci.";
RL J. Biol. Chem. 259:3686-3693(1984).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER M PROTEINS.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20374; AAA26976.1; -
CC PIR; A03501; MMSOMP.
CC PIR; A28616; A28616.
CC InterPro; IPR001899; -
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC PRINTS; PR00015; GPOSANCHOR.
CC PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
CC VIRULENCE; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
KW Transmembrane; Coiled coil; Signal.
FT SIGNAL 1 42
FT CHAIN 43 492 M PROTEIN, SEROTYPE 5.
FT DOMAIN 43 456 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 457 486 MEMBRANE ANCHOR.
FT DOMAIN 487 492 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 421 457 GLY/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 458 463 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
FT DOMAIN 69 103 5 X 7 AA TANDEM REPEATS OF L-K-T-K-N-E-G.
FT REPEAT 69 75 1.
FT REPEAT 76 82 2.
FT REPEAT 83 89 3.
FT REPEAT 90 96 4.
FT REPEAT 97 103 5.
FT CONFLICT 43 43 A -> T (IN REF. 2).
FT CONFLICT 50 50 N -> S (IN REF. 2).
FT CONFLICT 102 102 K -> SNLERKTAELTSEK (IN REF. 2).
FT CONFLICT 208 208 I -> L (IN REF. 2).
FT CONFLICT 492 492 87377986CBD55E27 CRC64;
SQ SEQUENCE 492 AA; 55085 MW; 87377986CBD55E27 CRC64;
Query Match 15.1%; Score 58; DB 1; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.4e-48;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 AVTRGTINDPQRAKALDKYELNHDLTKNGLKTEENGLKTEENGLKTE 123
DB 43 AVTRGTINDPQRAKALDKYELNHDLTKNGLKTEENGLKTEENGLKTE 100
RESULT 3
M6_STRPY
ID M6_STRPY PRT; 483 AA.
AC P08089;

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE M PROTEIN, SEROTYPE 6 PRECURSOR.
GN EMM6.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_taxid=1314;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE-86111835; PubMed-3511046;
RA Hollingshead S.K., Fischetti V.F., Scott J.R.;
RT "Complete nucleotide sequence of type 6 M protein of the group A
RT Streptococcus. Repetitive structure and membrane anchor.";
RL J. Biol. Chem. 261:1677-1686(1986).
RN [2]
RN SEQUENCE OF 43-122 FROM N.A.
RX MEDLINE-85166224; PubMed-3885219;
RA Scott J.R., Pulliam W.M., Hollingshead S.K., Fischetti V.A.;
RT "Relationship of M protein genes in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1822-1826(1985).
RN [3]
RN REVIEW.
RX MEDLINE-91126460; PubMed-1846974;
RA Fischetti V.A.;
RT "Neural vector. Herpes may open the way to gene therapy in neurons.";
RL Sci. Am. 264:32-39(1991).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER M PROTEINS.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11338; AAA26920.1; -
CC PIR; A26297; A26297.
CC InterPro; IPR001899; -
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC PRINTS; PR00015; GPOSANCHOR.
CC PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
CC VIRULENCE; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
KW Transmembrane; Coiled coil; Signal.
FT SIGNAL 1 42
FT CHAIN 43 483 M PROTEIN, SEROTYPE 6.
FT DOMAIN 43 457 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 458 477 MEMBRANE ANCHOR.
FT DOMAIN 478 483 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 69 138 10 X 7 AA TANDEM REPEATS.
FT DOMAIN 157 269 4.5 X 25 AA TANDEM REPEATS.
FT DOMAIN 279 347 TWO DIRECTLY REPEATED 27 AMINO ACID
FT PROTEINS. BLOCKS SEPARATED BY 15 AMINO ACIDS.
FT DOMAIN 348 411 HYDROPHILIC.
FT DOMAIN 412 448 GLY/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 449 454 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 483 AA; 53472 MW; 68F87F28DB53A448 CRC64;
Query Match 9.1%; Score 35; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 3.2e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Tue May 15 07:28:42 2001

us-09-151-409-16.oligo.rsp

Page 3

Oy 126 RVFPRGTVENPDKARELLNKYDVENSMLQANDKL 160
|||||
Db 43 RVFPRGTVENPDKARELLNKYDVENSMLQANDKL 77
|||||

Search completed: May 10, 2001, 08:43:09
Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:40:53 ; Search time 27.21 Seconds
(without alignments)
1649.784 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size : 10

Total number of hits satisfying chosen parameters: 24

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_15.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-unclassified.*
13: sp-vertebrate.*
14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	35.0	247	2 Q53344	Q53344 streptococc
2	52	13.6	457	2 Q54510	Q54510 streptococc
3	50	13.1	76	2 Q53536	Q53536 streptococc
4	50	13.1	79	2 Q54543	Q54543 streptococc
5	50	13.1	100	2 Q54632	Q54632 streptococc
6	50	13.1	272	2 Q9KGR9	Q9KGR9 streptococc
7	50	13.1	355	2 Q10371	Q10371 streptococc
8	50	13.1	484	2 Q10372	Q10372 streptococc
9	50	13.1	539	2 Q54719	Q54719 streptococc
10	45	11.7	100	2 Q54635	Q54635 streptococc
11	42	11.0	231	2 Q9KGR8	Q9KGR8 streptococc
12	38	9.9	100	2 Q54637	Q54637 streptococc
13	36	9.4	84	2 Q54598	Q54598 streptococc
14	35	9.1	99	2 Q54561	Q54561 streptococc
15	35	9.1	150	2 Q54587	Q54587 streptococc
16	30	7.8	488	2 Q54830	Q54830 streptococc
17	30	7.8	581	2 Q54835	Q54835 streptococc
18	26	6.8	100	2 Q54638	Q54638 streptococc
19	24	6.3	97	2 Q54845	Q54845 streptococc

20 23 6.0 83 2 086878 086878 streptococc
21 21 5.5 558 2 054718 054718 streptococc
22 19 5.0 83 2 Q92EH6 Q92EH6 streptococc
23 18 4.7 87 2 Q54548 Q54548 streptococc
24 14 3.7 87 2 Q54546 Q54546 streptococc

ALIGNMENTS

RESULT 1

Q53344 PRELIMINARY; PRT; 247 AA.
AC Q53344;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE TETRAVALENT M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93346765; PubMed=8345202;
RA Dale J.B., Chiang E.Y., Lederer J.W.;
RT "Recombinant tetraivalent group A streptococcal M protein vaccine.";
RL J. Immunol. 151:2188-2194(1993).
DR EMBL; S64396; AAB27745.1; -.
FT NON_TER 247
SQ SEQUENCE 247 AA; 28252 MW; 74A5827DE945529B CRC64;

Query Match 35.0%; Score 134; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 7.2e-122;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 GSAVTRGTINDPQRAKEALDKYLENHDLTKNEGLKTENEGLKTENEGLKTE 123
Db 114 GSAVTRGTINDPQRAKEALDKYLENHDLTKNEGLKTENEGLKTENEGLKTE 173
QY 124 VDRVFPRTGVENPKARELLNKYDVENSMLQANDKLPWRVRYTRHTPEDKLLKIIDDL 183
Db 174 VDRVFPRTGVENPKARELLNKYDVENSMLQANDKLPWRVRYTRHTPEDKLLKIIDDL 233
QY 184 AKEHELOQQNEKLS 197
Db 234 AKEHELOQQNEKLS 247

RESULT 2

Q54510 PRELIMINARY; PRT; 457 AA.
AC Q54510;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE M5.8193 PROTEIN (FRAGMENT).
GN EMM5.8193.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC8193;
RX MEDLINE=94224154; PubMed=8170398;
RA Whatmore A.M., Kehoe M.A.;
RT "Horizontal gene transfer in the evolution of group A streptococcal
emm-like genes: gene mosaics and variation in Vir regulons.";
RL Mol. Microbiol. 11:363-374(1994).
DR EMBL; U02480; AAA50854.1; -.

```
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 457 AA; 51172 MW; DEBBAF550AEB6379 CRC64;

Query Match 13.6%; Score 52; DB 2; Length 457;
Best Local Similarity 100.0%; Pred. No. 4.8e-42; Indels 0; Gaps 0;
Matches 52; Conservative 0; Mismatches 0;

Qy 67 VTRGTINDPQRAEKALDKYLENHDLKTKNEGLKTKTENEGKLTKTENEGKLTKTENE 118
Db 20 VTRGTINDPQRAEKALDKYLENHDLKTKNEGLKTKTENEGKLTKTENEGKLTKTENE 71

RESULT 3
ID Q53536 PRELIMINARY; PRT; 76 AA.
AC Q53536;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95377598; PubMed=7649435;
RA Penney T.J., Martin D.R., Williams L.C., de Malmanche S.A.,
RA Bergquist P.L.;
RT "A single emm gene-specific oligonucleotide probe does not recognise
RT all members of the Streptococcus pyogenes M type 1.";
RL FEMS Microbiol. Lett. 130:145-149(1995).
DR EMBL; S79254; AAB35162.1; -.
FT NON_TER 1
SQ SEQUENCE 76 AA; 8383 MW; EA46B64E5657A220 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 8.4e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPATQNTLRHKNKDLKARLENAMVAGRDFKRA 249
Db 22 NGDGNPREVIEDLAANNPATQNTLRHKNKDLKARLENAMVAGRDFKRA 71

RESULT 4
ID Q54543 PRELIMINARY; PRT; 79 AA.
AC Q54543;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE M TYPE 1 (EMML) (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 1.
RA Whatmore A.M.;
RL Thesis (1993), University of Newcastle Upon Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 1.
```

```
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
RT and the population genetic structure of group A streptococci.";
RL MOJ. Microbiol. 14:619-631(1994).
DR EMBL; U11940; AAA99556.1; -.
DR INTERPRO; IPR000886; -.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 79
SQ SEQUENCE 79 AA; 8499 MW; AE8A1FF8A4ELB711 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 8.7e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPATQNTLRHKNKDLKARLENAMVAGRDFKRA 249
Db 27 NGDGNPREVIEDLAANNPATQNTLRHKNKDLKARLENAMVAGRDFKRA 76

RESULT 5
ID Q54632 PRELIMINARY; PRT; 100 AA.
AC Q54632;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE M1.0 PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MGAS 2116;
RX MEDLINE=95172752; PubMed=7868273;
RA Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.;
RT "Genetic diversity and relationships among Streptococcus pyogenes
RT strains expressing serotype M1 protein: recent intercontinental spread
RT of a subclone causing episodes of invasive disease.";
RL Infect. Immun. 63:994-1003(1995).
DR EMBL; U20094; AAA85107.1; -.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11275 MW; 63316019FC2B6CF2 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.1e-40; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPATQNTLRHKNKDLKARLENAMVAGRDFKRA 249
Db 16 NGDGNPREVIEDLAANNPATQNTLRHKNKDLKARLENAMVAGRDFKRA 65

RESULT 6
ID Q9KGR9 PRELIMINARY; PRT; 272 AA.
AC Q9KGR9;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN=3148-00;
 RA Beall B.W.;
 RT #5' emm sequence emm3.2.*;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF288752; AAF91496.1;
 FT NON_TER 272 272
 FT NON_TER 272 272
 SQ SEQUENCE 272 AA; 30862 MW; 026PB243D345A35A CRC64;

Query Match 13.1%; Score 50; DB 2; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2.6e-40;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 LLDQVLTLYTKHNSYQQYNAQAGRLDLQKAEYKGLNDWAERLLQELN 301
 |||||
 Db 49 LLDQVLTLYTKHNSYQQYNAQAGRLDLQKAEYKGLNDWAERLLQELN 98

RESULT 7
 Q10371
 ID Q10371 PRELIMINARY; PRT; 355 AA.
 AC Q10371;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMELrel. 15, Last annotation update)
 DE M PROTEIN, SEROTYPE 1.0 PRECURSOR (FRAGMENTS).
 GN EMM1.0.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS130;
 RX MEDLINE=93360826; PubMed=8355619;
 RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
 RT "Nucleotide substitutions and small-scale insertion produce size and
 RT antigenic variation in group A streptococcal M1 protein.";
 RL Mol. Microbiol. 8:981-991(1993).
 RN [2]

CC SEQUENCE OF 1-227 FROM N.A.
 CC STRAIN=CS130;
 RA Haanes-Fritz E., Kraus W., Burdett V., Dale J.B., Beachey E.H.,
 RA Cleary P.;
 RA Nucleic Acids Res. 16:4667-4676(1993).
 CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 DR EMBL: X07933; CAA30757.1;
 DR INTERPRO: IPR001899;
 DR PRINTS: PR00015; GPOSANCHOR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
 KW Transmembrane; Coiled coil; Signal.

FT SIGNAL 1 41
 FT CHAIN 42 355 M1.0 PROTEIN.
 FT NON_CONS 227 228
 FT DOMAIN 42 329 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 134 227 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
 FT DOMAIN 284 320 GLY/PRO-RICH.
 FT DOMAIN 321 325 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.

FT TRANSMEM 330 349 MEMBRANE ANCHOR.
 FT DOMAIN 350 355 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 134 161 A-1.
 FT REPEAT 162 189 A-2.
 FT REPEAT 190 >227 A-3 (PARTIAL).
 SQ SEQUENCE 355 AA; 39595 MW; FF839CCAECD801B9 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 3.3e-40;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIQIRLRHENKDLKARLENAMVAGRDFKRA 249
 |||||
 Db 42 NGDGNPREVIEDLAANNPAIQIRLRHENKDLKARLENAMVAGRDFKRA 91

RESULT 8
 Q10372
 ID Q10372 PRELIMINARY; PRT; 484 AA.
 AC Q10372;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMELrel. 13, Last annotation update)
 DE M PROTEIN, SEROTYPE 1.2 PRECURSOR.
 GN EMM1.2.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=T1/29/58;
 RX MEDLINE=93360826; PubMed=8355619;
 RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
 RT "Nucleotide substitutions and small-scale insertion produce size and
 RT antigenic variation in group A streptococcal M1 protein.";
 RL Mol. Microbiol. 8:981-991(1993).

CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 DR EMBL: X62131; CAA44062.1;
 DR INTERPRO: IPR001899;
 DR PRAM: PF00746; Gram_pos_anchor; 1.
 DR PRINTS: PR00015; GPOSANCHOR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
 KW Transmembrane; Coiled coil; Signal.

FT SIGNAL 1 41
 FT CHAIN 42 484 M1.2 PROTEIN.
 FT DOMAIN 42 458 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 134 244 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
 FT DOMAIN 245 379 3 X APPROXIMATE TANDEM REPEATS, C-TYPE.
 FT DOMAIN 413 449 GLY/PRO-RICH.
 FT DOMAIN 450 455 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.

FT TRANSMEM 459 478 MEMBRANE ANCHOR.
 FT DOMAIN 479 484 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 134 161 A-1.
 FT REPEAT 162 189 A-2.
 FT REPEAT 190 244 A-3.
 FT REPEAT 245 286 C-1.
 FT REPEAT 287 327 C-2.
 FT REPEAT 328 379 C-3.
 SQ SEQUENCE 484 AA; 54234 MW; 4179A7BBAC5FAED9 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 484;
 Best Local Similarity 100.0%; Pred. No. 4.4e-40;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIQIRLRHENKDLKARLENAMVAGRDFKRA 249
 |||||
 Db 42 NGDGNPREVIEDLAANNPAIQIRLRHENKDLKARLENAMVAGRDFKRA 91

RESULT 9
 Q54719
 ID Q54719 PRELIMINARY; PRT; 539 AA.

```
AC Q54719;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE M3 PROTEIN.
GN EM3.1.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE 3;
RA Dale J.B., Simmons M., Chiang E., Chiang E.;
RL Vaccine 0:0-0(0).
DR EMBL; U0231; AAA96960.1; -.
DR HSSP; P03069; IGCL.
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 539 AA; 60163 MW; 93D20A4302C30F51 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 539;
Best Local Similarity 100.0%; Pred. No. 4.8e-40;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 LLDQVTLTKHNSNYQYNAQAGRLDLRQKAEYKLGINDWAERLLQELN 301
Db 63 LLDQVTLTKHNSNYQYNAQAGRLDLRQKAEYKLGINDWAERLLQELN 112

RESULT 10
Q54635
ID Q54635 PRELIMINARY; PRT; 100 AA.
AC Q54635;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE M1.3 PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS 249;
RX MEDLINE=95172752; PubMed=7868273;
RA Musser J.M., Kapur V., Szezo J., Pan X., Swanson D.S., Martin D.R.;
RT "Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease.";
RL Infect. Immun. 63:994-1003(1995).
DR EMBL; U20097; AAA85110.1; -.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11248 MW; 76FE8A3C33C15CF2 CRC64;

Query Match 11.7%; Score 45; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 7.7e-36;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 PREVEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 249
Db 21 PREVEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 65

RESULT 11
Q9KGR8
ID Q9KGR8 PRELIMINARY; PRT; 231 AA.
AC Q9KGR8;
```

```
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7203-99;
RA Beall B.W.;
RL "5' emm sequence emm3.3.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288753; AAF91497.1; -.
FT NON_TER 1
FT NON_TER 231
SQ SEQUENCE 231 AA; 26305 MW; F5002430DD229351 CRC64;

Query Match 11.0%; Score 42; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.3e-32;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 YTKHNSNYQYNAQAGRLDLRQKAEYKLGINDWAERLLQELN 301
Db 58 YTKHNSNYQYNAQAGRLDLRQKAEYKLGINDWAERLLQELN 99

RESULT 12
Q54637
ID Q54637 PRELIMINARY; PRT; 100 AA.
AC Q54637;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE M1.5 PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS 1666;
RX MEDLINE=95172752; PubMed=7868273;
RA Musser J.M., Kapur V., Szezo J., Pan X., Swanson D.S., Martin D.R.;
RT "Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease.";
RL Infect. Immun. 63:994-1003(1995).
DR EMBL; U20099; AAA85112.1; -.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11217 MW; 5332ACD6FFE7A3F2 CRC64;

Query Match 9.9%; Score 38; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.8e-29;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 LAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 249
Db 28 LAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 65

RESULT 13
Q54598
ID Q54598 PRELIMINARY; PRT; 84 AA.
AC Q54598;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
```



```
DE M TYPE 68 (EMML) (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 68;
RA Whatmore A.M.;
RL Thesis (1993), University of Newcastle Upon Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 68;
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci."
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11997; AAA99613.1; -.
FT NON_TER 1
FT NON_TER 84
SQ SEQUENCE 84 AA; 9123 MW; BE68CCD5AF3C518F CRC64;

Query Match 9.4%; Score 36; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 3.6e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 ANPAQNTLRRLHKNKDLKARLENAMVAGRDFKRA 249
      |||||
DB 45 ANPAQNTLRRLHKNKDLKARLENAMVAGRDFKRA 80

RESULT 14
Q54561 PRELIMINARY; PRT; 99 AA.
AC Q54561;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE M TYPE 19 (EMML) (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 19;
RA Whatmore A.M.;
RL Thesis (1993), University of Newcastle Upon Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 19;
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci."
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11959; AAA99575.1; -.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11034 MW; B04429D3DD2DF2E8 CRC64;

Query Match 9.1%; Score 35; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.8e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRYRTHPPEDKLLKIIDLDKAHEHLOQQONEKLK 197
      |||||
DB 31 RVRYRTHPPEDKLLKIIDLDKAHEHLOQQONEKLK 65
```

```
RESULT 15
Q54587 PRELIMINARY; PRT; 150 AA.
AC Q54587;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE EMML PROTEIN (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 6;
RA Whatmore A.M.;
RT "Sequence Analysis of the Emm-Like Gene Family of Streptococcus
Pyogenes."
RL Thesis (1993), Microbiology, University of Newcastle-Upon-Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=M TYPE 6;
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci."
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11986; AAA99602.1; -.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16572 MW; 8E00EAA28AC71157 CRC64;
```

Query Match 9.1%; Score 35; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 5.6e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
 |||||
DB 30 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 64

Search completed: May 10, 2001, 08:42:54
Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:57:50 ; Search time 1527.63 Seconds
(without alignments)
11179.422 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacgatggtcgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1283235 seqs, 7373929652 residues

Word size : 10
Total number of hits satisfying chosen parameters: 988154

Minimum DB seq length: 7
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl:

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_p11:*
- 13: gb_p12:*
- 14: gb_p13:*
- 15: gb_p14:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_higo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vi:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_v12:*
- 60: gb_v12:*
- 61: gb_v12:*
- 62: gb_v12:*
- 63: gb_v12:*
- 64: gb_v12:*
- 65: gb_v12:*
- 66: gb_v12:*
- 67: gb_v12:*
- 68: gb_v12:*
- 69: gb_v12:*
- 70: gb_v12:*
- 71: gb_v12:*
- 72: gb_v12:*
- 73: gb_v12:*
- 74: gb_v12:*
- 75: gb_v12:*
- 76: gb_v12:*
- 77: gb_v12:*
- 78: gb_v12:*
- 79: gb_v12:*
- 80: gb_v12:*
- 81: gb_v12:*
- 82: gb_v12:*
- 83: gb_v12:*
- 84: gb_v12:*
- 85: gb_v12:*
- 86: gb_v12:*
- 87: gb_v12:*
- 88: gb_v12:*
- 89: gb_v12:*
- 90: gb_v12:*
- 91: gb_v12:*
- 92: gb_v12:*
- 93: gb_v12:*
- 94: gb_v12:*
- 95: gb_v12:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	402	34.7	741	3	S64396
2	240	20.7	2131	3	STREMM24
3	174	15.0	1547	3	STRMPSA
4	158	13.6	2633	3	SPU02480
5	151	13.0	229	3	S79254
6	151	13.0	237	3	SPU11940
7	151	13.0	300	3	SPU20094
8	151	13.0	817	2	AF288752
9	151	13.0	1332	9	AR079002
10	151	13.0	1494	3	SPENM1M
11	151	13.0	1620	3	SPU40231

- S64396 emm=tetrava
- M19031 S.pyogenes
- M20374 S.pyogenes
- U02480 Streptococc
- S79254 emm=M prote
- U11940 Streptococc
- U20094 Streptococc
- AF288752 Streptococ
- AR079002 Sequence
- X62131 S.pyogenes
- U40231 Streptococc

C 888	17	1.5	164808	92	HSJ293L8	AL078594	Human DNA
C 889	17	1.5	164815	63	AP0013668	AL013668	Homo sapi
C 890	17	1.5	164839	13	AP002844	AP002844	Oryza sat
C 891	17	1.5	165175	66	AC019277	AC019277	Homo sapi
C 892	17	1.5	165405	61	AC009420	AC009420	Homo sapi
C 893	17	1.5	165436	92	HSJ160H22	AL121983	Human DNA
C 894	17	1.5	165454	89	AL138880	AL138880	Human DNA
C 895	17	1.5	165618	90	AL442636	AL442636	Human DNA
C 896	17	1.5	165622	83	CNS06C87	AL391397	Homo sapi
C 897	17	1.5	165663	90	AL445647	AL45647	Human DNA
C 898	17	1.5	166051	83	CNS06C71	AL389895	Homo sapi
C 899	17	1.5	166105	69	AC024972	AC024972	Homo sapi
C 900	17	1.5	166204	70	AC026763	AC026763	Homo sapi
C 901	17	1.5	166218	61	AC009293	AC009293	Homo sapi
C 902	17	1.5	166308	83	CNS01DX2	AL139020	Homo sapi
C 903	17	1.5	166321	65	AC018434	AC018434	Homo sapi
C 904	17	1.5	166335	75	AC073799	AC073799	Mus muscu
C 905	17	1.5	166357	88	AC026307	AC026307	Homo sapi
C 906	17	1.5	166487	83	CNS05TDS	AL357093	Homo sapi
C 907	17	1.5	166554	69	AC025296	AC025296	Oryza sat
C 908	17	1.5	166625	65	AC018523	AC018523	Homo sapi
C 909	17	1.5	166629	73	AC064847	AC064847	Homo sapi
C 910	17	1.5	166706	83	HSBA268E6	AL109803	Homo sapi
C 911	17	1.5	166728	70	AC025646	AC025646	Homo sapi
C 912	17	1.5	166868	91	CNS05TEP	AL359238	Human chr
C 913	17	1.5	166995	67	AC022631	AC022631	Homo sapi
C 914	17	1.5	167101	85	AC002094	AC002094	Genomic s
C 915	17	1.5	167182	88	AC069208	AC069208	Homo sapi
C 916	17	1.5	167205	73	AC067955	AC067955	Homo sapi
C 917	17	1.5	167275	75	AC074123	AC074123	Homo sapi
C 918	17	1.5	167348	63	AC013782	AC013782	Homo sapi
C 919	17	1.5	167357	69	AC025342	AC025342	Homo sapi
C 920	17	1.5	167467	84	AC069071	AC069071	Homo sapi
C 921	17	1.5	167584	72	AP001843	AP001843	Homo sapi
C 922	17	1.5	167607	79	AL162394	AL162394	Homo sapi
C 923	17	1.5	167623	75	AC073450	AC073450	Homo sapi
C 924	17	1.5	167655	68	AC023192	AC023192	Homo sapi
C 925	17	1.5	167764	71	AC027523	AC027523	Homo sapi
C 926	17	1.5	168068	90	CNS05TFC6	AL355072	Human chr
C 927	17	1.5	168111	70	AC025788	AC025788	Homo sapi
C 928	17	1.5	168153	88	AF127577	AF127577	Homo sapi
C 929	17	1.5	168258	65	AC018703	AC018703	Homo sapi
C 930	17	1.5	168259	73	AC067842	AC067842	Homo sapi
C 931	17	1.5	168344	83	AP003062	AP003062	Homo sapi
C 932	17	1.5	168383	82	AP000676	AP000676	Homo sapi
C 933	17	1.5	168654	88	AC064878	AC064878	Homo sapi
C 934	17	1.5	168720	76	AC078781	AC078781	Homo sapi
C 935	17	1.5	168805	74	AC069518	AC069518	Homo sapi
C 936	17	1.5	168855	70	AC028258	AC028258	Homo sapi
C 937	17	1.5	169039	77	AC084741	AC084741	Homo sapi
C 938	17	1.5	169148	75	AC073023	AC073023	Homo sapi
C 939	17	1.5	169235	80	AL358373	AL358373	Homo sapi
C 940	17	1.5	169241	81	AL445987	AL445987	Homo sapi
C 941	17	1.5	169400	67	AC022208	AC022208	Homo sapi
C 942	17	1.5	169543	65	AC016811	AC016811	Homo sapi
C 943	17	1.5	169598	75	AC073140	AC073140	Homo sapi
C 944	17	1.5	169599	70	AC026041	AC026041	Homo sapi
C 945	17	1.5	169603	80	AL356134	AL356134	Homo sapi
C 946	17	1.5	169660	87	AC011479	AC011479	Homo sapi
C 947	17	1.5	169752	87	AC021953	AC021953	Homo sapi
C 948	17	1.5	169812	65	AC016845	AC016845	Homo sapi
C 949	17	1.5	169813	72	AC055735	AC055735	Homo sapi
C 950	17	1.5	169892	65	AC018512	AC018512	Homo sapi
C 951	17	1.5	169939	76	AC079765	AC079765	Homo sapi
C 952	17	1.5	170026	63	AC013627	AC013627	Homo sapi
C 953	17	1.5	170067	83	CNS01RGI	AL157997	Homo sapi
C 954	17	1.5	170195	80	AL357556	AL357556	Homo sapi
C 955	17	1.5	170291	79	AL353607	AL353607	Homo sapi
C 956	17	1.5	170402	13	AP002070	AP002070	Oryza sat
C 957	17	1.5	170546	76	AC074296	AC074296	Homo sapi
C 958	17	1.5	170578	77	AC079887	AC079887	Oryza sat
C 959	17	1.5	170610	68	AC023935	AC023935	Homo sapi
C 960	17	1.5	170630	72	AC040990	AC040990	Homo sapi

ALIGNMENTS

RESULT	1	S64396	741 bp	DNA	15-OCT-1993
LOCUS		S64396			
DEFINITION		emm=tetravalent M protein [Streptococcus, group A, Genomic Recombinant, 741 nt].			
ACCESSION		S64396			
VERSION		S64396.1	GI:408224		
KEYWORDS		Streptococcus group A.			
SOURCE		Streptococcus			
ORGANISM		Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae.			
REFERENCE		1 (bases 1 to 741)			
AUTHORS		Dale,J.B., Chiang,E.Y. and Lederer,J.W			
TITLE		Recombinant tetravalent group A streptococcal M protein vaccine			
JOURNAL		J. Immunol. 151 (4), 2188-2194 (1993)			
MEDLINE		93346765			
REMARK		GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 136070] from the original journal article. This sequence comes from Fig. 1.			
FEATURES		Location/Qualifiers			
source		1..741			
		/organism="Streptococcus"			
		/db_xref="taxon:1301"			
gene		1..741			
		/partial			
		/gene="emm"			
		/note="tetravalent M protein"			
CDS		1..741			

```

/organism="Streptococcus pyogenes"
/db_xref="taxon:1314"
157. .1776
/note="type 24 M-protein precursor"
/codon_start=1
/transl_table=11
/protein_id="AAA26874.1"
/db_xref="GI:153617"
/translation="MTKNNTNRHYSLRKLKTGTA
TRSDYTLLEKVOERADKFETENRNLTKNSDSL
KNDKLSLKASKAKIOELERAKDEKALEGAMNFS
EKALBEGAMNFSSTADSAKITLEAKAALBARQAE
EKALAALAKDEKALEGAMNFSSTADSAKITLEA
TADSAKITLEAKAALBARQAEKDEKALHQSQV
LEEONKIKSEASQSLRDLDSREAKKQLEAEHQ
EAKKQVKALEEANSKLALEKLNKLEESKRLTI
QAEELALIKRAGKSDSQTPDAPKGNKAVPGQAI
GETANPFFTAALTVMATAGVAAVVKKEN"
157. .282
/note="pot. type 24 M-protein signal"
283. .1773
/note="type 24 M-protein"
1802. .1838
/misc_feature
848 a 376 c 421 g 486 t
100 bp upstream of HincII site.
BASE COUNT
ORIGIN

```

	Query Match	20.7%	Score 240;	DB 3;	Length 2131;
	Best Local Similarity	100.0%;	Pred. No. 2.4e-117;	Mismatches 0;	Indels 0;
	Matches 240;	Conservative 0;	Gaps 0;		
Qy	910	gtcgcgactagdtctcagacagatactctgaaaaagtacaagaacgctcgacaagttt	969		
Dd	283	GTCCGCGACTAGTCTCAGACAGATACCTGGANAAAGTCACGAACGTCGTGCACAAGTTT	342		
Qy	970	gagatagaaaaacaatacgtttaaaacttaagaatatagtgacttaagttttaataataaaacgcg	1029		
Dd	343	GAGATAGAAAACAATACTGTTAAACTTAAGAATATAGTGACTTAAGTTTAAATATAAAGCG	402		
Qy	1030	ttaaagatcataatgatgatgttaactgaagagttgagtaatgctaaagagaaactacgt	1089		
Dd	403	TTAAAGATCATATGATGAGTTAACTGAACAGTTGAGTAATGCTAAACAGCAACTACGT	462		
Qy	1090	aaaatcatataaacatctctgaaaagactagtaaaaattcaagaatttagaggcacgtaag	1149		
Dd	463	AAAAATGATNAATCACCTATCTGAAAAGCTAGTAAAAATTCAAGAAATTAGAGGCGACGTAAAG	522		

```

3
RESULT
STRSMP5A      STRSMP5A      1547 bp      DNA      BCT      26-APR-1993
LOCUS          S.pyogenes smp5 gene encoding serotype 5 M protein, complete cds.
DEFINITION     M20374
ACCESSION      M20374.1 GI:153812
VERSION         5 M protein; smp5 gene.
KEYWORDS       S.pyogenes (strain Manfredo) DNA, clone pMK207.
SOURCE         Streptococcus pyogenes
ORGANISM       Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
                Streptococcus.
REFERENCE      1 (bases 1 to 1547)
AUTHORS       Miller, L., Gray, L., Beachey, E.H. and Kehoe, M.
TITLE          Antigenic variation among group A streptococcal M proteins:
                Nucleotide sequence of the serotype 5 M protein gene and its
                relationship with genes encoding types 6 and 24 M proteins
                J. Biol. Chem. 263, 5668-5673 (1988)
JOURNAL        88186881
MEDLINE
FEATURES       Location/Qualifiers
                1..1547
                /organism="Streptococcus pyogenes"
                /db_xref="taxon:1314"
                43..1521
                /note="5 M protein"
SOURCE
CDS

```

```

/codon_start=1
/transl_table=11
/protein_id="AAA26976.1"
/db_xref="GI:153813"
/translation="MARENTHKIYWLRLKKGTSYVAVALSVLGAGLVYVNTSEVSAV
TRGTINDPQRAKEALDEKLENTENHDLTKTKNEGLKTENEGLTENEGLTKEGLTKK
EHEAENDKLLQOORLLTSTQKETLEREVONTQYNNETLUKIKNGDLTKELNKTROELAK
QOESKEKKALELLELTVDKDTAKQENKETTGLKKILDETIVKOKIAKEQNKETI
GTLKKILDETIVKDLKAEQSKQKNGIKALQKDLDAKANKISDASRGLRRDLDASRE
AKKOLEHDTQKLEQNKISPASRKLRRDLDAKAKQLEAQOQKLEQNKISSEAR
KGLRRDLDAKAKQKVEKALEEANSLLAEKLNKLEESKLTKEKKAELQAKLEA
EAKALQELQAKQAEALAKLRAGKASDSQTPDTFGNKAVPKGCOATPOAGCTKPNQKAP
MKETKRLQPSGTENAPFFTAALTYVATAGVAAVARKEEN"
238..258
/repeat_region
/note="repeat type A"
259..279
/repeat_region
/note="repeat type A"
280..300
/repeat_region
/note="repeat type A"
301..321
/repeat_region
/note="repeat type A"
322..342
/repeat_region
/note="repeat type A"
343..363
/repeat_region
/note="repeat type A"
364..384
/repeat_region
/note="repeat type A"
385..405
/repeat_region
/note="repeat type B"
406..426
/repeat_region
/note="repeat type B"
427..447
/repeat_region
/note="repeat type B"
448..468
/repeat_region
/note="repeat type B"
469..489
/repeat_region
/note="repeat type B"
490..510
/repeat_region
/note="repeat type B"
511..531
/repeat_region
/note="repeat type B"
532..552
/repeat_region
/note="repeat type B"
553..573
/repeat_region
/note="repeat type B"
574..594
/repeat_region
/note="repeat type B"
595..615
/repeat_region
/note="repeat type B"
616..636
/repeat_region
/note="repeat type B"
637..657
/repeat_region
/note="repeat type B"
658..678
/repeat_region
/note="repeat type B"
679..699
/repeat_region
/note="repeat type B"
700..720
/repeat_region
/note="repeat type B"
721..741
/repeat_region
/note="repeat type B"
742..762
/repeat_region
/note="repeat type B"
763..783
/repeat_region
/note="repeat type B"
784..804
/repeat_region
/note="repeat type B"
805..825
/repeat_region
/note="repeat type B"
826..846
/repeat_region
/note="repeat type B"
847..867
/repeat_region
/note="repeat type B"
868..888
/repeat_region
/note="repeat type B"
889..909
/repeat_region
/note="repeat type B"
910..930
/repeat_region
/note="repeat type B"
931..951
/repeat_region
/note="repeat type B"
952..972
/repeat_region
/note="repeat type B"
973..993
/repeat_region
/note="repeat type B"
994..1014
/repeat_region
/note="repeat type B"
1015..1035
/repeat_region
/note="repeat type B"
1036..1056
/repeat_region
/note="repeat type B"
1057..1077
/repeat_region
/note="repeat type B"
1078..1098
/repeat_region
/note="repeat type B"
1099..1119
/repeat_region
/note="repeat type B"
1120..1140
/repeat_region
/note="repeat type B"
1141..1161
/repeat_region
/note="repeat type B"
1162..1182
/repeat_region
/note="repeat type B"
1183..1203
/repeat_region
/note="repeat type B"
1204..1224
/repeat_region
/note="repeat type B"
1225..1245
/repeat_region
/note="repeat type B"
1246..1266
/repeat_region
/note="repeat type B"
1267..1287
/repeat_region
/note="repeat type B"
1288..1308
/repeat_region
/note="repeat type B"
1309..1329
/repeat_region
/note="repeat type B"
1330..1350
/repeat_region
/note="repeat type B"
1351..1371
/repeat_region
/note="repeat type B"
1372..1392
/repeat_region
/note="repeat type B"
1393..1413
/repeat_region
/note="repeat type B"
1414..1434
/repeat_region
/note="repeat type B"
1435..1455
/repeat_region
/note="repeat type B"
1456..1476
/repeat_region
/note="repeat type B"
1477..1497
/repeat_region
/note="repeat type B"
1498..1518
/repeat_region
/note="repeat type B"
1519..1539
/repeat_region
/note="repeat type B"
1540..1560
/repeat_region
/note="repeat type B"
1561..1581
/repeat_region
/note="repeat type B"
1582..1602
/repeat_region
/note="repeat type B"
1603..1623
/repeat_region
/note="repeat type B"
1624..1644
/repeat_region
/note="repeat type B"
1645..1665
/repeat_region
/note="repeat type B"
1666..1686
/repeat_region
/note="repeat type B"
1687..1707
/repeat_region
/note="repeat type B"
1708..1728
/repeat_region
/note="repeat type B"
1729..1749
/repeat_region
/note="repeat type B"
1750..1770
/repeat_region
/note="repeat type B"
1771..1791
/repeat_region
/note="repeat type B"
1792..1812
/repeat_region
/note="repeat type B"
1813..1833
/repeat_region
/note="repeat type B"
1834..1854
/repeat_region
/note="repeat type B"
1855..1875
/repeat_region
/note="repeat type B"
1876..1896
/repeat_region
/note="repeat type B"
1897..1917
/repeat_region
/note="repeat type B"
1918..1938
/repeat_region
/note="repeat type B"
1939..1959
/repeat_region
/note="repeat type B"
1960..1980
/repeat_region
/note="repeat type B"
1981..2001
/repeat_region
/note="repeat type B"
2002..2022
/repeat_region
/note="repeat type B"
2023..2043
/repeat_region
/note="repeat type B"
2044..2064
/repeat_region
/note="repeat type B"
2065..2085
/repeat_region
/note="repeat type B"
2086..2106
/repeat_region
/note="repeat type B"
2107..2127
/repeat_region
/note="repeat type B"
2128..2148
/repeat_region
/note="repeat type B"
2149..2169
/repeat_region
/note="repeat type B"
2170..2190
/repeat_region
/note="repeat type B"
2191..2211
/repeat_region
/note="repeat type B"
2212..2232
/repeat_region
/note="repeat type B"
2233..2253
/repeat_region
/note="repeat type B"
2254..2274
/repeat_region
/note="repeat type B"
2275..2295
/repeat_region
/note="repeat type B"
2296..2316
/repeat_region
/note="repeat type B"
2317..2337
/repeat_region
/note="repeat type B"
2338..2358
/repeat_region
/note="repeat type B"
2359..2379
/repeat_region
/note="repeat type B"
2380..2400
/repeat_region
/note="repeat type B"
2401..2421
/repeat_region
/note="repeat type B"
2422..2442
/repeat_region
/note="repeat type B"
2443..2463
/repeat_region
/note="repeat type B"
2464..2484
/repeat_region
/note="repeat type B"
2485..2505
/repeat_region
/note="repeat type B"
2506..2526
/repeat_region
/note="repeat type B"
2527..2547
/repeat_region
/note="repeat type B"
2548..2568
/repeat_region
/note="repeat type B"
2569..2589
/repeat_region
/note="repeat type B"
2590..2610
/repeat_region
/note="repeat type B"
2611..2631
/repeat_region
/note="repeat type B"
2632..2652
/repeat_region
/note="repeat type B"
2653..2673
/repeat_region
/note="repeat type B"
2674..2694
/repeat_region
/note="repeat type B"
2695..2715
/repeat_region
/note="repeat type B"
2716..2736
/repeat_region
/note="repeat type B"
2737..2757
/repeat_region
/note="repeat type B"
2758..2778
/repeat_region
/note="repeat type B"
2779..2799
/repeat_region
/note="repeat type B"
2800..2820
/repeat_region
/note="repeat type B"
2821..2841
/repeat_region
/note="repeat type B"
2842..2862
/repeat_region
/note="repeat type B"
2863..2883
/repeat_region
/note="repeat type B"
2884..2904
/repeat_region
/note="repeat type B"
2905..2925
/repeat_region
/note="repeat type B"
2926..2946
/repeat_region
/note="repeat type B"
2947..2967
/repeat_region
/note="repeat type B"
2968..2988
/repeat_region
/note="repeat type B"
2989..3009
/repeat_region
/note="repeat type B"
3010..3030
/repeat_region
/note="repeat type B"
3031..3051
/repeat_region
/note="repeat type B"
3052..3072
/repeat_region
/note="repeat type B"
3073..3093
/repeat_region
/note="repeat type B"
3094..3114
/repeat_region
/note="repeat type B"
3115..3135
/repeat_region
/note="repeat type B"
3136..3156
/repeat_region
/note="repeat type B"
3157..3177
/repeat_region
/note="repeat type B"
3178..3198
/repeat_region
/note="repeat type B"
3199..3219
/repeat_region
/note="repeat type B"
3220..3240
/repeat_region
/note="repeat type B"
3241..3261
/repeat_region
/note="repeat type B"
3262..3282
/repeat_region
/note="repeat type B"
3283..3303
/repeat_region
/note="repeat type B"
3304..3324
/repeat_region
/note="repeat type B"
3325..3345
/repeat_region
/note="repeat type B"
3346..3366
/repeat_region
/note="repeat type B"
3367..3387
/repeat_region
/note="repeat type B"
3388..3408
/repeat_region
/note="repeat type B"
3409..3429
/repeat_region
/note="repeat type B"
3430..3450
/repeat_region
/note="repeat type B"
3451..3471
/repeat_region
/note="repeat type B"
3472..3492
/repeat_region
/note="repeat type B"
3493..3513
/repeat_region
/note="repeat type B"
3514..3534
/repeat_region
/note="repeat type B"
3535..3555
/repeat_region
/note="repeat type B"
3556..3576
/repeat_region
/note="repeat type B"
3577..3597
/repeat_region
/note="repeat type B"
3598..3618
/repeat_region
/note="repeat type B"
3619..3639
/repeat_region
/note="repeat type B"
3640..3660
/repeat_region
/note="repeat type B"
3661..3681
/repeat_region
/note="repeat type B"
3682..3702
/repeat_region
/note="repeat type B"
3703..3723
/repeat_region
/note="repeat type B"
3724..3744
/repeat_region
/note="repeat type B"
3745..3764
/repeat_region
/note="repeat type B"
3765..3785
/repeat_region
/note="repeat type B"
3786..3806
/repeat_region
/note="repeat type B"
3807..3827
/repeat_region
/note="repeat type B"
3828..3848
/repeat_region
/note="repeat type B"
3849..3869
/repeat_region
/note="repeat type B"
3870..3890

```

```

Qy 317 taaaactgagaatgaagggttaaaactgagaatgaa 354
Db 176 TAAAACTGAGATGAGGGTTAAAAAAGTGAATGAA 213

RESULT 5
S79254
LOCUS S79254 229-bp DNA BCT 30-NOV-1995
DEFINITION emm-M protein [Streptococcus pyogenes, M1a, genomic, 229 nt].
ACCESSION S79254
VERSION S79254.1 GI:1087104
KEYWORDS Streptococcus pyogenes M1a.
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 229)
AUTHORS Penney,T.J., Martin,D.R., Williams,L.C., de Malmarche,S.A. and Bergquist,P.L.
TITLE A single emm gene-specific oligonucleotide probe does not recognise all members of the Streptococcus pyogenes M type 1
JOURNAL FEMS Microbiol. Lett. 130 (2-3), 145-149 (1995)
MEDLINE 95377598
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 169934] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
source
1..229
Location/Qualifiers
/organism="Streptococcus pyogenes"
/db_xref="taxon:1314"
2..229
/gene="emm"
partial
/feature="emm"
/feature="M protein"
2..229
/gene="emm"
partial
/feature="emm"
/feature="M protein"
note="This sequence comes from Fig. 2"
/codon_start=1
/product="M protein"
/protein_id="AAB35162.1"
/db_xref="GI:1087105"
/translation="SVAVRLTVLCAGFANOTEVKANGDNPREVIEDLAANNPAIQNI
RLRHENKDLKARLENMEVAGRDFKRAELEK"
BASE COUNT 84 a 32 c 59 g 54 t
ORIGIN

Query Match 13.0%; Score 151; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 598 aacggtgatgtaacctaggaagttagaagattcgcagcaaaacaaatcccgcaata 657
Db 65 AACGGTGATGTTAATCCTAGGGAAGTTATAGAAGATCTTGACGACAAACATCCCGCAATA 124

Qy 658 caaaatatacgtttacgtcacgaaacaagagacttaaaagcgagattagagaatgcaatg 717
Db 125 CAAAATATACGTTTACGTCCAGCAAAACAGGACTTAAAGCGAGATTAGAGAAATGCAATG 184

Qy 718 gaagttgcaggaagagattttaagagagctg 748
Db 185 GAAGTTGCAGGAGAGATTTTAAGAGAGCTG 215

RESULT 6
SPU11940
LOCUS SPU11940 237 bp DNA BCT 16-MAY-1996
DEFINITION Streptococcus pyogenes M type 1 (emm1) gene, partial cds.
ACCESSION U11940
VERSION U11940.1 GI:533557
KEYWORDS

SOURCE Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 300)
AUTHORS Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and Martin,D.R.
TITLE Genetic diversity and relationships among Streptococcus pyogenes

SOURCE Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 237)
AUTHORS Whatmore,A.M.
TITLE Sequence Analysis of the Emm-Like Gene Family of Streptococcus pyogenes
JOURNAL Thesis (1993) Microbiology, University of Newcastle Upon Tyne
AUTHORS Whatmore,A.M., Kapur,V., Sullivan,D.J., Musser,J.M. and Kehoe,M.A.
TITLE Non-congruent relationships between variation in emm gene sequences and the population genetic structure of group A streptococci
JOURNAL Mol. Microbiol. 14 (4), 619-631 (1994)
MEDLINE 95198537
REFERENCE 3 (bases 1 to 237)
AUTHORS Whatmore,A.M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1994) Embl Data Library By: A. M. Whatmore, Microbiology, University of Newcastle Upon Tyne, Framlington Place, Newcastle Upon Tyne, UK, NE2 4HH
FEATURES
source
1..237
Location/Qualifiers
/organism="Streptococcus pyogenes"
/strain="M type 1"
/isolate="NCTC8198"
/db_xref="taxon:1314"
1..237
/gene="emm1"
<1..>237
/gene="emm1"
/codon_start=1
/transl_table=11
/protein_id="AAA99556.1"
/db_xref="GI:533558"
/translation="KTGTASVAVALTVLGAFFANOTEVKANGDNPREVIEDLAANNP
AIONIRLHENKDLKARLENMEVAGRDFKRAEEL"
BASE COUNT 86 a 35 c 62 g 54 t
ORIGIN

Query Match 13.0%; Score 151; DB 3; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 598 aacggtgatgtaacctaggaagttagaagattcgcagcaaaacaaatcccgcaata 657
Db 79 AACGGTGATGTTAATCCTAGGGAAGTTATAGAAGATCTTGACGACAAACATCCCGCAATA 138

Qy 658 caaaatatacgtttacgtcacgaaacaagagacttaaaagcgagattagagaatgcaatg 717
Db 139 CAAAATATACGTTTACGTCCAGCAAAACAGGACTTAAAGCGAGATTAGAGAAATGCAATG 198

Qy 718 gaagttgcaggaagagattttaagagagctg 748
Db 199 GAAGTTGCAGGAGAGATTTTAAGAGAGCTG 229

RESULT 7
SPU20094
LOCUS SPU20094 300 bp DNA BCT 05-JAN-1996
DEFINITION Streptococcus pyogenes M1.0 protein gene, partial cds.
ACCESSION U20094
VERSION U20094.1 GI:643545
KEYWORDS Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 300)
AUTHORS Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and Martin,D.R.
TITLE Genetic diversity and relationships among Streptococcus pyogenes

```

strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease
Infect. Immun. 63 (3), 994-1003 (1995)
951/2752
REFERENCE 2 (bases 1 to 300)
AUTHORS Swanson, D.S.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COLLEGE OF MEDICINE, ONE BAYLOR PLAZA, HOUSTON, TX 77030, USA
FEATURES
source
1. .300
/organism="Streptococcus pyogenes"
/isolate="MGAS 2116"
/db_xref="taxon:1314"
<1..>300
/codon_start=1
/transl_table=11
/product="M1.0 protein"
/protein_id="AA085107.1"
/db_xref="GI:643546"
/translation="TVLGGAFANQTEVKANGDNPREVIEDLAANNPAIGNIRLRHEN
KDLKARLENAMVAGRDFKRAEELEKAKQALEQDKLETKLKLQDYDLAKESI"
KDLKARLENAMVAGRDFKRAEELEKAKQALEQDKLETKLKLQDYDLAKESI"
BASE COUNT 126 a 45 c 67 g 62 t
ORIGIN
Query Match 13.0%; Score 151; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.2e-69; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacggtgagtggaatccttaggaaagttagaagaatcttgcagcaaaacaatcccgcaata 657
DB 46 AACCGTGATGGTAACTTAGGCAAGTTATAGAAGATCTTGCAAGCAACAATCCGCAATA 105
QY 658 caaataatcgtttacgtcacaagaacagagcacttaagaagcaatagagaatgcaatg 717
DB 106 CAAATATACGTTTACGTCACCAACCAAGGACTTAAAGCGAGATTAGAGATGCAATG 165
QY 718 gaagttcgaggaagagattttaagagagctg 748
DB 166 GAAGTTGAGGAAGAGATTTTAAGAGAGCTG 196
RESULT 8
AF288752 817 bp DNA BCT 02-AUG-2000
LOCUS Streptococcus pyogenes isolate 3148-00 M protein (emm) gene,
partial cds.
DEFINITION
ACCESSION AF288752
VERSION AF288752.1 GI:9652381
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 817)
AUTHORS Beall, B.W.
TITLE 5' emm sequence emm3.2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 817)
AUTHORS Beall, B.W.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2000) Respiratory Diseases Branch, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA 30333, USA
FEATURES
source
1. .817
/organism="Streptococcus pyogenes"
/isolate="3148-00"
/db_xref="taxon:1314"
/note="blood isolate recovered from patient in New York population based surveillance; r non typeable; opacity factor negative"

group: A"
<1..>817
/gene="emm"
<1..>817
/gene="emm"
/note="emm3.2"
/codon_start=1
/transl_table=11
/product="M protein"
/protein_id="AAF91496.1"
/db_xref="GI:9652382"
/translation="TRSYAVALTVLGTGLVAGQTVKADARSARSYNGEPRHVKLKN
EIEILLDQVTLTKHNSYQYNAQAGRLDLRQKAEYTKGLNDAERLLGELNGEDV
KKVLGKVAPEKDDLEKKEKEDYQDLKDFDLAKGQYVLSDRHQOELE
EKERKVTATARKVQISBELETVKQKVEKSTMODLTEKQNRYSQLEQLATTKQNAKED
PELAALANAADKQKLEAKIADLETKEAKEDFELAALGHQHAHNEYQAKLAEK"
BASE COUNT 342 a 105 c 182 g 188 t
ORIGIN
Query Match 13.0%; Score 151; DB 2; Length 817;
Best Local Similarity 100.0%; Pred. No. 1.2e-69; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 753 cttgttagatcagggttacacaattatataataataataataataataataataa 812
DB 144 CTTGTTAGATCAGGTTACACAATTATATATAACATAATAAGTAAATACCAACATATAA 203
QY 813 tgcaacagctggcagacttgacctgagacaaaaggctgaatattctaaaggccttaata 872
DB 204 TGCACAAGCTGGCAGACTTGACCTGAGACAAAAGGCTGAATATCTAAAGGCTTAAATGA 263
QY 873 ttgggctgagagcgtgttacaaagagtgtaaat 903
DB 264 TTGGCTGAGAGGCTGTACAGAGTAAAT 294
RESULT 9
AR079002 1332 bp DNA PAT 31-AUG-2000
LOCUS Sequence 5 from patent US 5965390.
DEFINITION
ACCESSION AR079002
VERSION AR079002.1 GI:10005748
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1332)
AUTHORS Bjorck, L. and Sjobring, U.
TITLE Protein L and hybrid proteins thereof
JOURNAL Patent: US 5965390-A 5 12-OCT-1999;
FEATURES
source
1. .1332
/organism="unknown"
BASE COUNT 568 a 244 c 281 g 239 t
ORIGIN
Query Match 13.0%; Score 151; DB 9; Length 1332;
Best Local Similarity 100.0%; Pred. No. 1.2e-69; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacggtgagtggaatccttaggaaagttagaagaatcttgcagcaaaacaatcccgcaata 657
DB 1 AACCGTGATGGTAACTTAGGCAAGTTATAGAAGATCTTGCAAGCAACAATCCGCAATA 60
QY 658 caaataatcgtttacgtcacaagaacagagcacttaagaagcaatagagaatgcaatg 717
DB 61 CAAATATACGTTTACGTCACCAACCAAGGACTTAAAGCGAGATTAGAGATGCAATG 120
QY 718 gaagttcgaggaagagattttaagagagctg 748
DB 121 GAAGTTGAGGAAGAGATTTTAAGAGAGCTG 151

Qy	718	gaagtgcagaaagagattttaacagaactg	748
Dd	274	GAAAGTCGACGAAGAGATTTTAAGAGACTG	304
RESULT	11		
SPU40231			
LOCUS		1620 bp DNA BCT	15-APR-1996
DEFINITION		Streptococcus pyogenes M3 protein (emm3.1) gene, complete cds.	
ACCESSION		U40231	
VERSION		U40231.1 GI:1263022	
KEYWORDS			
SOURCE		Streptococcus pyogenes strain=Type 3.	
ORGANISM		Streptococcus pyogenes Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
REFERENCE			
AUTHORS		1. (bases 1 to 1620)	
TITLE		Dale,J.B., Simmons,M., Chiang,E. and Chiang,E.	
JOURNAL		Octavalent group A streptococcal M protein vaccine	
REFERENCE		Vaccine (1996) In press	
AUTHORS		2. (bases 1 to 1620)	
TITLE		Dale,J.B.	
JOURNAL		Direct Submission	
FEATURES		Submitted (06-NOV-1995) James B. Dale, Medicine, VA Medical Center, 1030 Jefferson Ave., Memphis, TN 38104, USA	
source		Location/Qualifiers	
gene		1..1620	
CDS		/organism="Streptococcus pyogenes" /strain="Type 3" /db_xref="taxon:1314"	
		1..1620	
		/gene="emm3.1"	
		1..1620	
		/gene="emm3.1"	
		/codon_start=1	
		/transl_table=11	
		/product="M3 protein"	
		/protein_id="AA96960.1"	
		/db_xref="GI:1263023"	
		/translation="MAKNNTNRHSYLRKLTGTASVAVALTVLGLTGLVAGQTVKADAR SYNGEPRIHVILKLKNEJLENLDQVTOLTYTHNSNYOYNQAQRGLDLRLROAEYLKGLD WNERLLQELNGEDVKKVLGVKAPEKDDEKEVEKIDKKEKYODLDDKDFDLAK GYVLSDRHQOELEEREKKVTEATAKVGOI SELETVKOKVESTMODLTEKQNRVSOL EOELATTKNQKEDFELALANAADQKLEAKIADLETCLKKEAFELALGHQHAK NEYQAKLAKSDQIKOLESQQLLDASRKGTDPHVPRAOKKATEAELNNLKAEHLAK VTEQKIOLASRGRTARDLEAVRQAQAVEALKLEQNRTSEASKGLRRDLDSAR EAKQVKEALEANSKLALEEKLESKLTEKEALQAKLEAEAKALKQEOLAK QAEELAKRAGASQSIPDTKPGNKAVPGKQPAQGTPKPNONKAPMKETTKRPST GETANPFTFAALTVMNATGAVAVVRKKEEN"	
BASE COUNT		697 a 269 c 335 g 319 t	
ORIGIN			
Query Match 13.0%; Score 151; DB 3; Length 1620;			
Best Local Similarity 100.0%; Pred. No. 1.2e-69;			
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	753	ctgttagatcagggttacacaattatactaacaataatagtaattaccacaataataa	812
Dd	186	CTTGTTAGTACAGTTACACATTATATATAACAATAATAGTAATTACCACAATAATA	245
QY	813	tgcacaagctggcagacttgacctgagacaaaaggctggaatatctaaaaagcccttaatga	872
Dd	246	TGCACAAGCTGGCAGACTTGACCTGAGACAAAAGCGTGAATATCTAAAGCCCTTAATGA	305
QY	873	ttggcgctgaaggctgttacaagagttaaat	903
Dd	306	TTGGGCTGAGAGGCTGTTTTCAAGAAGTTAAAT	336
RESULT	12		
SPUPM1			

LOCUS SPMM1 1708 bp DNA BCT 14-MAR-1996
DEFINITION S.pyogenes emm1 gene for type 1 group A M protein.
ACCESSION X07933 X07860
VERSION X07933.1 GI:47358
KEYWORDS emm1 gene; M protein.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1708)
AUTHORS Hannes-Fritz,E.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-1988) Hannes-Fritz E., University of Minnesota,
Department of Microbiology, Box 196, Mayo Building, Minneapolis, MN
55455
REFERENCE 2 (bases 1 to 1708)
AUTHORS Haanes-Fritz,E., Kraus,W., Burdett,V., Dale,J.B., Beachey,E.H. and
Cleary,P.
TITLE Comparison of the leader sequences of four group A streptococcal M
protein genes
JOURNAL Nucleic Acids Res. 16 (10), 4667-4677 (1988)
MEDLINE 88247768
FEATURES Location/Qualifiers
source
1..1708
/organism="Streptococcus pyogenes"
/strain="CS130"
/db_xref="taxon:1314"
/clones="pbf82"
921..963
/note="pot.promoter region"
1015..1019
1028..1708
/codon_start=1
/transl_table=11
/product="M protein precursor"
/protein_id="CAA30757.1"
/db_xref="GI:47359"
/db_xref="SPTREMBL:Q10371"
/translation="MAKNNTNRHYSRLKTKGTASVAVALTVLGAGFANTEVKANGD
GNPREVIDLAANNPAIONIRLHENKDLKARLENAMVAGRODFKRAEELEKAKQALE
DQRDLTKLEKQDDYDLAKESTSWDRQLEKELEKKEALELAIQASRDYHRATA
LEKELEKKKALELAIQASQDYNRANVLEKELETITREQEINRLNLGNKLELDQLS
SEKQLTIR"
sig_peptide 1028..1150
mat_peptide 1151..1708
repeat_region 1427..1525
/product="mature M protein, type 1 group A"
/rpt_type=DIRECT
repeat_region 1522..1609
/rpt_type=DIRECT
BASE COUNT 640 a 310 c 331 g 427 t
ORIGIN
1028..1150
1151..1708
/product="mature M protein, type 1 group A"
/rpt_type=DIRECT
1427..1525
/rpt_type=DIRECT
1522..1609
/rpt_type=DIRECT
Query Match 13.0%; Score 151; DB 3; Length 1708;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacgggtgagtgtaactcctaggaagttagagaagatcttgacgacaaacaatcccgcaata 657
|||||
DB 1151 AACGGTCATGGTAATCTAGGAGAGTTATAGAGATCTTGCGACGAACAATCCCGCAATA 1210
QY 658 caaaatatacgtttacgtccacgaacagagacttaaaagcgagattagagaatgcaatg 717
|||||
DB 1211 CAAAATATACGTTTAGGTACGACGAACAGGACTTAAAGCGGAGATTAGAGAATGCAATG 1270
QY 718 gaagttgcggaagaagatttaagagagctg 748
|||||
DB 1271 GAAGTTGCAGGAAGAGATTTAAGAGAGCTG 1301
RESULT 13
SPU20097

LOCUS SPU20097 300 bp DNA BCT 05-JAN-1996
DEFINITION Streptococcus pyogenes M1.3 protein gene, partial cds.
ACCESSION U20097
VERSION U20097.1 GI:643551
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 300)
AUTHORS Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and
Martin,D.R.
TITLE Genetic diversity and relationships among Streptococcus pyogenes
strains expressing serotype M1 protein: recent intercontinental
spread of a subclone causing episodes of invasive disease
Infect. Immun. 63 (3), 994-1003 (1995)
JOURNAL 95172752
MEDLINE 2 (bases 1 to 300)
REFERENCE Swanson,D.S.
AUTHORS Direct Submission
TITLE Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL
FEATURES Location/Qualifiers
source
1..300
/organism="Streptococcus pyogenes"
/isolate="MGAS 249"
/db_xref="taxon:1314"
<1..>300
/codon_start=1
/transl_table=11
/product="M1.3 protein"
/protein_id="AAH85110.1"
/db_xref="GI:643552"
/translation="TVLGAGFANTEVKANGDSPREVIEDLAANNPAIONIRLHEN
KDLKARLENAMVAGRODFKRAEELEKAKQALEDKRLTKLEKQDDYDLAKEST"
BASE COUNT 125 a 45 c 58 g 62 t
ORIGIN
11.8%; Score 137; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.9e-62;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 612 tcttagggaggttatagaagatcttgcagcaaacatccgcgaatacaaaatatacgttt 671
|||||
DB 60 TCTTAGGAAAGTTATAGAGATCTTGCGACGAACAATCCCGCAATACAAATATACGTTT 119
QY 672 acgtcacgaaaaaaggaacttaaaagcgagattagagaatgcaatggaattgcaggaag 731
|||||
DB 120 ACGTCACGAACAAACAGGACTTAAAGCGAGATTAGAGAAATGCAATGGAAAGTTGCAGGAAG 179
QY 732 agattttaagagagctg 748
|||||
DB 180 AGATTTTAAGAGAGCTG 196
RESULT 14
AF288753
LOCUS AF288753 694 bp DNA BCT 02-AUG-2000
DEFINITION Streptococcus pyogenes isolate 7203-99 M protein (emm) gene,
partial cds.
ACCESSION AF288753
VERSION AF288753.1 GI:9652383
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 694)
AUTHORS Beall,B.W.
TITLE 5' emm sequence emm3.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 694)


```
AUTHORS      Beall,B.W.
TITLE        Direct Submission
JOURNAL      Submitted (21-JUL-2000) Respiratory Diseases Branch, Centers for
              Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA
              30333, USA
FEATURES
Source
  Location/Qualifiers
    1..694
      /organism="Streptococcus pyogenes"
      /isolate="7203-99"
      /db_xref="taxon:1314"
      /note="type 3; opacity factor negative; blood isolate"
      group: A"
      <1..>694
      /gene="emm"
      <1..>694
      /gene="emm"
      /gene="emm3.3"
      /codon_start=1
      /transl_table=11
      /product="M protein"
      /protein_id="AAF91497.1"
      /db_xref="GI:9652384"
      /translation="TASVAVALTVLGTLVAGQTVKADARSVNGEFPRHVKLNKEIEN
      LLDVOTLYTKHIVTKHNSVQYQNAQGRDLDRKAEYLGKGLNDWAEERLQELNGED
      VKVLGKVAPEKFDLEKEVKELEKIDKKEKEYODLDKDFDLAKQGVLSDKRHQQEL
      EEKPKVTEATKVGQISEELETVKQVESTMDLTEKQNRVSOLEQELATTQNAKE
      DFLAALANAADK"
BASE COUNT   289 a   87 c   152 g   166 t
ORIGIN
Query Match      11.1%; Score 128; DB 2; Length 694;
Best Local Similarity 100.0%; Pred. No. 2.6e-57;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 776 tatatacaataatagtaattaccacaataataatgcacaaactgacgacttgacc 835
Db 170 TATATACTAAACATAATAGTATTACCAACATATATAGTCACAGCTGGCAGACTTGACC 229
QY 836 ttagacaaaggctgaatctctaaaggccttaagtattgggctgagaggctgtacaag 895
Db 230 TGAGACAAAAGCGTGAATATCTAAAGGCTTAATGATTGGGCTGAGAGGCTGTACAAG 289
QY 896 agttaaat 903
      |||||
Db 290 AGTTAAAT 297

RESULT 15
SPU20099
LOCUS      SPU20099      300 bp      DNA      BCT      05-JAN-1996
DEFINITION Streptococcus pyogenes M1.5 protein gene, partial cds.
ACCESSION  U20099
VERSION     U20099.1  GI:643555
KEYWORDS
SOURCE
ORGANISM   Streptococcus pyogenes.
            Streptococcus pyogenes
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE  1 (bases 1 to 300)
AUTHORS   Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and
            Martin,D.R.
TITLE     Genetic diversity and relationships among Streptococcus pyogenes
            strains expressing serotype M1 protein: recent intercontinental
            spread of a subclone causing episodes of invasive disease
            Infect. Immun. 63 (3), 994-1003 (1995)
            95172752
REFERENCE  2 (bases 1 to 300)
AUTHORS   Swanson,D.S.
TITLE     Direct Submission
JOURNAL   Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
FEATURES
Location/Qualifiers
```

```
source
1..300
/organism="Streptococcus pyogenes"
/isolate="MGAS 1666"
/db_xref="taxon:1314"
<1..>300
/codon_start=1
/transl_table=11
/product="M1.5 protein"
/protein_id="AAA85112.1"
/db_xref="GI:643556"
/translation="TVLGAGFANQTEVKANGDGNPREVIEGLAANPAIONIRLRHEN
KDLKARLENAMEVAGRDFKRAELEKAKALEDKRKLETKELQDDYDLAKEST"
BASE COUNT   125 a   45 c   68 g   62 t
ORIGIN
Query Match      10.0%; Score 116; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 7e-51;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 633 tcttgacgacaaacaatcccgaataacaaaataatacgtttacgtcacgaaacaaaggactt 692
Db 81 TCTTGCAGCAACAAATCCCGCAATACAAAATATACGTTTACGTACCAAAACAAAGGACTT 140
QY 693 aaaagcgagattagagagaatgcaatggaagttgcaggagaagagattttaagagagctg 748
Db 141 AAAAGCGAGATTAGAGAATGCAATGGAAGTTGCAGGAAGAGAGATTTTAAAGAGAGCTG 196
```

Search completed: May 13, 2001, 06:13:28
Job time: 4538 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:59:10 ; Search time 121.03 Seconds
(without alignments)
5585.547 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagatggtcgactag.....aggcagtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 678276 seqs, 291890651 residues

Word size : 10

Total number of hits satisfying chosen parameters: 214117

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1158	100.0	1158	20 X33103	Streptococcal hexa
2	402	34.7	741	21 A57894	S. pyogenes hybrid
3	351	30.3	741	15 Q45215	Recombinant M24-M5
4	240	20.7	861	15 Q45214	Recombinant M24-M5
5	240	20.7	861	21 A57893	Recombinant M24-M5
6	217	18.7	522	15 Q45220	S. pyogenes hybrid
7	217	18.7	522	21 A57899	Recombinant M24-M5
8	217	18.7	741	15 Q45219	S. pyogenes hybrid
9	217	18.7	741	21 A57898	Recombinant M19-M6
10	217	18.7	822	15 Q45216	S. pyogenes hybrid
11	217	18.7	822	21 A57895	Recombinant M24-M5
					S. pyogenes hybrid

12	217	18.7	1029	15 Q45221	Tetravalent-C repe
13	217	18.7	1029	21 A57900	S. pyogenes hybrid
14	105	9.1	1617	18 T49317	Type-6 M-protein e
15	63	5.4	301	6 M50341	Sequence of a port
16	53	4.6	561	15 Q45217	Recombinant M24-M5
17	53	4.6	561	21 A57896	S. pyogenes hybrid
18	52	4.5	204	15 Q45222	M24-M5-M6-M19 hybr
19	52	4.5	204	21 A57901	S. pyogenes hybrid
20	52	4.5	408	15 Q45223	S. pyogenes hybrid
21	52	4.5	408	21 A57902	M24-M5-M6-M19-M3-M
22	52	4.5	918	15 Q45218	S. pyogenes octava
23	52	4.5	918	21 A57897	Recombinant M24-M5
24	48	4.1	765	15 Q45160	S. pyogenes hybrid
25	48	4.1	855	15 Q45161	Sequence of a frag
26	38	3.3	39	21 A57905	Sequence of a frag
27	36	3.1	417	15 Q45159	Streptococcus pyog
28	34	2.9	47	21 A57907	Sequence of LT-B-M
29	32	2.8	54	21 A57917	Streptococcus pyog
30	27	2.3	33	20 X33102	Streptococcus pyog
31	25	2.2	30	15 Q45210	Streptococcal hexa
32	25	2.2	30	20 X30302	M6 top strand prim
33	25	2.2	30	21 A57913	Streptococcal hexa
34	25	2.2	36	20 X30298	Streptococcal hexa
35	24	2.1	30	15 Q45208	Streptococcal hexa
36	24	2.1	30	15 Q45209	M5 bottom strand p
37	24	2.1	30	15 Q45212	M9 top strand pri
38	24	2.1	30	15 Q45213	M9 bottom strand
39	24	2.1	30	20 X33099	Streptococcal hexa
40	24	2.1	30	20 X33100	Streptococcal hexa
41	24	2.1	30	20 X33101	Streptococcal hexa
42	24	2.1	30	20 X30299	Streptococcal hexa
43	24	2.1	30	20 X30300	Streptococcal hexa
44	24	2.1	30	20 X30301	Streptococcal hexa
45	24	2.1	30	20 X30303	Streptococcal hexa
46	24	2.1	30	20 X30304	Streptococcal hexa
47	24	2.1	30	20 X30305	Streptococcal hexa
48	24	2.1	30	20 X30306	Streptococcal hexa
49	24	2.1	30	20 X33098	Streptococcal hexa
50	24	2.1	30	21 A57911	Streptococcal hexa
51	24	2.1	30	21 A57915	Streptococcal hexa
52	24	2.1	30	21 A57916	Streptococcus pyog
53	22	1.9	29	15 Q45206	Streptococcus pyog
54	22	1.9	29	21 A57903	Streptococcus pyog
55	22	1.9	32	20 X03263	M24 top strand pri
56	22	1.9	32	20 X03263	Streptococcus pyog
57	19	1.6	390	16 T23376	Streptococcus pyog
58	19	1.6	390	21 A57918	Streptococcus pyog
59	19	1.6	2367	21 C42916	Human gene signatu
60	18	1.6	784	18 T49484	Human gene signatu
61	18	1.6	1260	21 C64427	Arabidopsis thalia
62	18	1.6	1392	18 T72874	Staphylococcus aur
63	18	1.6	5836	20 X86366	Staphylococcus aur
64	18	1.6	10336	20 X34652	DNA encoding glycy
65	17	1.5	623	21 C35048	DNA encoding the h
66	17	1.5	719	21 F12228	Wheat starch solub
67	17	1.5	994	21 C34777	Arabidopsis thalia
68	17	1.5	995	21 C35114	Aspergillus oryzae
69	17	1.5	1090	20 X13612	Arabidopsis thalia
70	17	1.5	1109	21 C54025	Arabidopsis thalia
71	17	1.5	1120	21 C59916	Enterococcus faeca
72	17	1.5	1440	21 Z37035	Arabidopsis thalia
73	17	1.5	1866	19 V53502	Human secreted pro
74	17	1.5	2858	18 T62650	DNA of the agAB ge
75	17	1.5	2858	18 T62650	DNA encoding a sta
76	17	1.5	2858	18 T64391	DNA encoding OCHI
77	17	1.5	2858	18 T59328	Pichia pastoris al
78	17	1.5	3096	8 N71063	OCHI gene. Pichia
79	17	1.5	3929	21 F21072	Genomic sequence e
80	17	1.5	3929	21 A34950	Human low adenosin
81	17	1.5	4668	19 V60918	Human adenosine re
82	17	1.5	5077	16 Q97812	Angiotensin gene.
83	17	1.5	5480	20 X13069	Sequence encoding
84	17	1.5	5912	21 C75806	Enterococcus faeca
			9834	18 V74348	Human ORFX ORF1361
					Staphylococcus aur

85	17	1.5	10011	19	V52209	Streptococcus pneu	158	16	1.4	1520	18	T80564	Human plasma plate
86	17	1.5	12447	21	F21073	Human low adenosin	159	16	1.4	1520	18	T63701	cDNA encoding plat
87	17	1.5	12447	21	A34951	Human adenosine re	160	16	1.4	1520	19	T96127	Human plasma plate
88	17	1.5	21636	21	A55966	Human G713 3'-end	161	16	1.4	1520	20	X08463	Human plasma plate
89	17	1.5	151826	21	F22291	BAC containing rep	162	16	1.4	1520	20	V08534	Human PAF-AH codin
90	16	1.4	161	21	Z29588	Rat cDNA clone rC7	163	16	1.4	1520	21	A59579	cDNA encoding plas
91	16	1.4	196	21	C23382	Human secreted pro	164	16	1.4	1520	21	A10861	Human platelet-act
92	16	1.4	272	21	C31922	Human secreted pro	165	16	1.4	1520	21	A22420	Human PAF-AH cDNA
93	16	1.4	273	21	C30896	Human secreted pro	166	16	1.4	1520	22	C89057	Platelet-activatin
94	16	1.4	283	21	C71545	Single nucleotide	167	16	1.4	1520	22	V16524	DNA encoding a Bac
95	16	1.4	296	21	C03603	Human secreted pro	168	16	1.4	1533	19	X83885	Bacillus thuringie
96	16	1.4	313	21	C03027	Human secreted pro	169	16	1.4	1549	20	C41964	Arabidopsis thalia
97	16	1.4	313	21	C03066	Human secreted pro	170	16	1.4	1685	21	A26702	Candida albicans p
98	16	1.4	314	21	C03056	Human secreted pro	171	16	1.4	1770	20	X35569	Yeast sphingosine
99	16	1.4	342	17	T30755	Mouse cryptdin 17	172	16	1.4	1808	21	A51405	Chromosome 16q tum
100	16	1.4	352	17	T30749	Mouse cryptdin 11	173	16	1.4	1918	21	C36056	Arabidopsis thalia
101	16	1.4	375	17	T30748	Mouse cryptdin 10	174	16	1.4	1923	21	C47422	Arabidopsis thalia
102	16	1.4	377	17	T30747	Mouse cryptdin 9 c	175	16	1.4	1925	21	C38450	Human lipolysis st
103	16	1.4	378	21	C04281	Human secreted pro	176	16	1.4	1934	20	X83422	Human lipolysis st
104	16	1.4	395	21	C23689	Human secreted pro	177	16	1.4	1954	21	A50287	C. felis esterase,
105	16	1.4	401	17	T30753	Mouse cryptdin 15	178	16	1.4	1982	19	V40735	C. felis esterase,
106	16	1.4	419	17	T30731	Mouse cryptdin 6 g	179	16	1.4	1982	19	V40736	Human lipolysis st
107	16	1.4	419	17	T30732	Mouse inactivated	180	16	1.4	2101	20	X83421	Human lipolysis st
108	16	1.4	419	17	T30727	Mouse cryptdin 1 g	181	16	1.4	2101	20	A50286	Corn HS2 storage p
109	16	1.4	422	17	T30739	Mouse cryptdin 1 c	182	16	1.4	2123	13	T034797	Corn seed storage
110	16	1.4	422	17	T30744	Mouse cryptdin 6 c	183	16	1.4	2123	17	T06488	Human 7TM clone HD
111	16	1.4	464	21	F12724	Aspergillus oryzae	184	16	1.4	2136	22	C82671	C. felis esterase,
112	16	1.4	472	21	C53594	Arabidopsis thalia	185	16	1.4	2144	19	V40758	C. felis esterase,
113	16	1.4	494	21	F12709	Aspergillus oryzae	186	16	1.4	2144	19	V40759	Human lipolysis st
114	16	1.4	511	21	A69825	Human ovarian carc	187	16	1.4	2158	20	X83420	Human lipolysis st
115	16	1.4	516	21	A79643	Pinus radiata cell	188	16	1.4	2158	21	A50285	Arabidopsis thalia
116	16	1.4	528	15	O64171	Mouse PAP. Mus mu	189	16	1.4	2217	21	C40913	Candida albicans e
117	16	1.4	536	18	X30810	Streptococcus pneu	190	16	1.4	2283	21	A52791	DNA encoding a Bac
118	16	1.4	577	16	O87960	Human acetyl hydro	191	16	1.4	2366	19	V16529	Bacillus thuringie
119	16	1.4	577	16	T87059	Human platelet-act	192	16	1.4	2366	20	X83890	DNA encoding a Bac
120	16	1.4	577	18	T80575	Human plasma plate	193	16	1.4	2369	19	V16526	DNA encoding a Bac
121	16	1.4	577	18	T66132	PAF-AH exon 8 and	194	16	1.4	2369	20	X83887	Bacillus thuringie
122	16	1.4	577	19	T96142	Human plasma plate	195	16	1.4	2375	19	V16521	DNA encoding a Bac
123	16	1.4	577	20	X08471	Exon 16 of human p	196	16	1.4	2375	19	V16522	DNA encoding a Bac
124	16	1.4	577	20	X08545	Human PAF-AH codin	197	16	1.4	2375	20	X83882	Bacillus thuringie
125	16	1.4	577	20	A59587	Plasma platelet-ac	198	16	1.4	2375	20	X83883	Bacillus thuringie
126	16	1.4	577	21	A10872	Human PAF-AH exon	199	16	1.4	2538	21	A51351	Bacillus thuringie
127	16	1.4	577	21	Z24248	Human PAF-AH exon	200	16	1.4	2545	19	V42980	Soybean sucrose no
128	16	1.4	577	22	C89065	Platelet-activatin	201	16	1.4	2774	15	O66579	Streptococcus pneu
129	16	1.4	623	19	T98674	DNA encoding a S.	202	16	1.4	2807	21	A11389	Sorbitol dehydroge
130	16	1.4	675	18	C37604	Arabidopsis thalia	203	16	1.4	2840	21	C78596	Mucor circinelloid
131	16	1.4	675	18	V75198	Staphylococcus aur	204	16	1.4	2840	19	V66452	Human PROM4993 nucl
132	16	1.4	688	21	C79807	Human secreted pro	205	16	1.4	2996	20	X35295	An anticdisease gen
133	16	1.4	702	10	N90852	Intron 3 from huma	206	16	1.4	2996	20	X35283	Nucleic acid encod
134	16	1.4	744	18	V75277	Staphylococcus aur	207	16	1.4	2996	21	A07263	Nucleic acid encod
135	16	1.4	756	19	V35234	Mouse FLS-ob/ob CD	208	16	1.4	2996	21	A07264	Rice probenazole re
136	16	1.4	809	21	Z60388	A diacylglycerol a	209	16	1.4	3160	22	A89290	Rice probenazole re
137	16	1.4	814	22	F22448	Human breast cance	210	16	1.4	3303	21	A70147	Drosophila site 1
138	16	1.4	862	21	C40425	Arabidopsis thalia	211	16	1.4	3303	21	F20910	Human defensin 3 p
139	16	1.4	907	21	C32722	Arabidopsis thalia	212	16	1.4	3710	21	A34788	Human adenosine re
140	16	1.4	984	20	X75836	H. pylori outer me	213	16	1.4	3865	18	T96833	Intron 3 of human
141	16	1.4	1060	21	Z29143	Rice phytoene synt	214	16	1.4	3865	18	T96833	Human cell surface
142	16	1.4	1077	21	A64808	C. trachomatis clo	215	16	1.4	3969	19	V05762	Chicken transmembr
143	16	1.4	1181	21	C60051	Human secreted pro	216	16	1.4	3969	20	V81897	Streptococcus pneu
144	16	1.4	1181	21	C37425	Arabidopsis thalia	217	16	1.4	5305	19	V52268	C. albicans gene C
145	16	1.4	1240	21	C52338	Arabidopsis thalia	218	16	1.4	5344	21	A10735	Mutant BRCA1 allel
146	16	1.4	1312	21	C34756	Arabidopsis thalia	219	16	1.4	5556	17	T42021	BRCA1 allele #5803
147	16	1.4	1335	18	T87072	Human platelet-act	220	16	1.4	5556	18	T70084	Mutant BRCA1 allel
148	16	1.4	1335	21	A10885	Synthetic PAF-AH c	221	16	1.4	5656	19	V60560	Tumourogenic BRCA1
149	16	1.4	1335	21	Z24261	Human PAF-AH cDNA	222	16	1.4	5656	18	T42023	Mutant BRCA1 allel
150	16	1.4	1335	22	C89078	Platelet-activatin	223	16	1.4	5689	17	T70066	Tumourogenic BRCA1
151	16	1.4	1361	16	Q81780	Human T-cell lymph	224	16	1.4	5689	17	V60562	BRCA1 allele #582
152	16	1.4	1361	21	A52357	cDNA encoding huma	225	16	1.4	5707	17	T42031	Mutant BRCA1 allel
153	16	1.4	1486	21	C51465	Arabidopsis thalia	226	16	1.4	5707	18	T70074	Tumourogenic BRCA1
154	16	1.4	1488	21	C33887	Arabidopsis thalia	227	16	1.4	5709	17	V60570	BRCA1 allele #4401
155	16	1.4	1503	21	C77783	Human cancer assoc	228	16	1.4	5709	17	T42027	BRCA1 allele #4406
156	16	1.4	1520	16	Q87947	Human platelet act	229	16	1.4	5709	17	T42028	BRCA1 allele #1020
157	16	1.4	1520	18	T87048	Human platelet-act	230	16	1.4	5709	17	T42029	

231	1.4	5709	17	T42022	BRCA1 allele #9601	304	1.4	5913	17	T17463	Mutated BRCA1 codi
232	1.4	5709	18	T70065	Mutant BRCA1 allele	305	1.4	5913	17	T17465	Mutated BRCA1 codi
233	1.4	5709	18	T70072	Mutant BRCA1 allele	306	1.4	5913	17	T17467	Mutated BRCA1 codi
234	1.4	5709	18	T70071	Mutant BRCA1 allele	307	1.4	5913	17	T17470	Mutated BRCA1 codi
235	1.4	5709	18	T70070	Mutant BRCA1 allele	308	1.4	5913	17	T17477	Mutated BRCA1 codi
236	1.4	5709	19	V60561	Tumorigenic BRCA1	309	1.4	5913	17	T17440	Mutated BRCA1 codi
237	1.4	5709	19	V60566	Tumorigenic BRCA1	310	1.4	5914	17	T32601	BRCA1, breast and
238	1.4	5709	19	V60567	Tumorigenic BRCA1	311	1.4	5914	17	T17445	Mutated BRCA1 codi
239	1.4	5709	19	V60568	Tumorigenic BRCA1	312	1.4	5914	17	T17446	Mutated BRCA1 codi
240	1.4	5709	19	V11738	Human BRCA1 DNA 38	313	1.4	5914	17	T17449	Mutated BRCA1 codi
241	1.4	5710	17	T42026	BRCA1 allele #388	314	1.4	5914	17	T17453	Mutated BRCA1 codi
242	1.4	5710	18	T70069	Mutant BRCA1 allele	315	1.4	5914	17	T17457	Mutated BRCA1 codi
243	1.4	5710	19	V60565	Tumorigenic BRCA1	316	1.4	5914	17	T17462	Mutated BRCA1 codi
244	1.4	5711	17	T42024	BRCA1 allele #8403	317	1.4	5914	17	T17466	Mutated BRCA1 codi
245	1.4	5711	17	T42030	BRCA1 allele #7408	318	1.4	5914	17	T17468	Mutated BRCA1 codi
246	1.4	5711	18	T87085	Human BRCA1 gene c	319	1.4	5914	17	T17471	Mutated BRCA1 codi
247	1.4	5711	18	T70067	Mutant BRCA1 allele	320	1.4	5914	17	T17476	Mutated BRCA1 codi
248	1.4	5711	18	T70073	Mutant BRCA1 allele	321	1.4	5914	17	T17478	Mutated BRCA1 codi
249	1.4	5711	19	V62180	BRCA1 (om1) codin	322	1.4	5914	17	T17479	Mutated BRCA1 codi
250	1.4	5711	19	V60563	Tumorigenic BRCA1	323	1.4	5914	17	T17481	Mutated BRCA1 codi
251	1.4	5711	19	V60569	Tumorigenic BRCA1	324	1.4	5914	17	T17483	Mutated BRCA1 codi
252	1.4	5711	19	V46463	Human BRCA1 omi2 p	325	1.4	5914	17	T17484	Mutated BRCA1 codi
253	1.4	5711	19	V46463	Human BRCA1 omi2 p	326	1.4	5914	17	T17486	Mutated BRCA1 codi
254	1.4	5711	19	V46465	Human BRCA1 omi3 p	327	1.4	5914	17	T17489	Mutated BRCA1 codi
255	1.4	5711	19	V46466	Human BRCA1 omi3 p	328	1.4	5914	17	T17490	Mutated BRCA1 codi
256	1.4	5711	19	V46466	Human BRCA1 omi3 p	329	1.4	5914	17	T17491	Mutated BRCA1 codi
257	1.4	5711	19	V46467	Human BRCA1 omi3 p	330	1.4	5914	17	T17492	Mutated BRCA1 codi
258	1.4	5711	19	V46468	Human BRCA1 omi3 p	331	1.4	5914	17	T17493	Mutated BRCA1 codi
259	1.4	5711	19	V46469	Human BRCA1 omi3 p	332	1.4	5914	17	T17494	Mutated BRCA1 codi
260	1.4	5711	19	V46470	Human BRCA1 omi3 p	333	1.4	5914	17	T17495	Mutated BRCA1 codi
261	1.4	5711	19	V46471	Human BRCA1 omi3 p	334	1.4	5914	17	T17496	Mutated BRCA1 codi
262	1.4	5711	19	V46449	Human BRCA1 omi2 c	335	1.4	5914	17	T17497	Mutated BRCA1 codi
263	1.4	5711	19	V46450	Human BRCA1 omi3 c	336	1.4	5914	17	T17498	Mutated BRCA1 codi
264	1.4	5711	19	V46451	Human BRCA1 omi1 p	337	1.4	5914	17	T17499	Mutated BRCA1 codi
265	1.4	5711	19	V46452	Human BRCA1 omi1 p	338	1.4	5914	17	T17500	Mutated BRCA1 codi
266	1.4	5711	19	V46453	Human BRCA1 omi1 p	339	1.4	5914	17	T17501	Mutated BRCA1 codi
267	1.4	5711	19	V46454	Human BRCA1 omi1 p	340	1.4	5914	17	T17502	Mutated BRCA1 codi
268	1.4	5711	19	V46455	Human BRCA1 omi1 p	341	1.4	5914	17	T17504	Mutated BRCA1 codi
269	1.4	5711	19	V46456	Human BRCA1 omi1 p	342	1.4	5914	17	T17505	Mutated BRCA1 codi
270	1.4	5711	19	V46457	Human BRCA1 omi1 p	343	1.4	5914	17	T17506	Mutated BRCA1 codi
271	1.4	5711	19	V46458	Human BRCA1 omi1 p	344	1.4	5914	17	T17507	Mutated BRCA1 codi
272	1.4	5711	19	V46459	Human BRCA1 omi2 p	345	1.4	5914	17	T17508	Mutated BRCA1 codi
273	1.4	5711	19	V46460	Human BRCA1 omi2 p	346	1.4	5914	17	T17509	Mutated BRCA1 codi
274	1.4	5711	19	V46461	Human BRCA1 omi2 p	347	1.4	5914	17	T17510	Mutated BRCA1 codi
275	1.4	5711	19	V46462	Human BRCA1 omi2 p	348	1.4	5914	17	T17511	Mutated BRCA1 codi
276	1.4	5711	19	V46463	Human BRCA1 omi1 c	349	1.4	5914	17	T17438	BRCA1 coding seque
277	1.4	5711	19	V11737	Human BRCA1 DNA	350	1.4	5914	17	T17439	Mutated BRCA1 codi
278	1.4	5711	21	C60793	Human BRCA1 (om1)	351	1.4	5914	17	T17441	Mutated BRCA1 codi
279	1.4	5711	21	C60794	Human BRCA1 (om2)	352	1.4	5914	17	T17442	Mutated BRCA1 codi
280	1.4	5711	21	C60795	Human BRCA1 (om3)	353	1.4	5914	17	T17444	Mutated BRCA1 codi
281	1.4	5712	16	Q93027	Gene encoding BRCA	354	1.4	5914	17	T18310	BRCA1, breast and
282	1.4	5712	17	T42032	BRCA1 allele #77	355	1.4	5914	17	T18311	BRCA1 gene nonsens
283	1.4	5712	18	T84840	Human breast and o	356	1.4	5914	17	T18313	BRCA1 gene missens
284	1.4	5712	18	T70075	Mutant BRCA1 allele	357	1.4	5915	17	T17451	Mutated BRCA1 codi
285	1.4	5712	19	V60571	Tumorigenic BRCA1	358	1.4	5915	17	T17458	Mutated BRCA1 codi
286	1.4	5712	21	T87995	BRCA1 gene sequenc	359	1.4	5915	17	T17460	Mutated BRCA1 codi
287	1.4	5770	18	T70068	Mutant BRCA1 allele	360	1.4	5915	17	T17461	Mutated BRCA1 codi
288	1.4	5770	19	V60564	Tumorigenic BRCA1	361	1.4	5915	17	T17488	Mutated BRCA1 codi
289	1.4	5874	17	T17459	Mutated BRCA1 codi	362	1.4	5915	17	T18312	BRCA1 gene framesh
290	1.4	5903	17	T17487	Mutated BRCA1 codi	363	1.4	5919	17	T17472	Mutated BRCA1 codi
291	1.4	5903	17	T18314	Mutated BRCA1 codi	364	1.4	5924	17	T17476	Mutated BRCA1 codi
292	1.4	5910	17	T17448	BRCA1 gene framesh	365	1.4	5940	21	A70105	Mutated BRCA1 codi
293	1.4	5910	17	T17469	Mutated BRCA1 codi	366	1.4	6578	17	T18767	Plasmodium falcipa
294	1.4	5910	17	T17473	Mutated BRCA1 codi	367	1.4	6578	19	X18171	RSV RNA-dependent
295	1.4	5910	17	T17474	Mutated BRCA1 codi	368	1.4	6578	19	X43109	Respiratory syncyt
296	1.4	5910	17	T17475	Mutated BRCA1 codi	369	1.4	8084	20	X13109	Enterococcus faeca
297	1.4	5910	17	T17480	Mutated BRCA1 codi	370	1.4	8365	21	E21445	Human defensin 2 p
298	1.4	5910	17	T17482	Mutated BRCA1 codi	371	1.4	9351	20	Z22301	cDNA encoding a hu
299	1.4	5912	17	T17447	Mutated BRCA1 codi	372	1.4	13086	18	V74327	Staphylococcus aur
300	1.4	5912	17	T17452	Mutated BRCA1 codi	373	1.4	15016	20	X59560	Nucleic acid seque
301	1.4	5912	17	T17464	Mutated BRCA1 codi	374	1.4	15210	20	X59703	Polynucleotide seq
302	1.4	5912	17	T17485	Mutated BRCA1 codi	375	1.4	15223	18	T78440	Human respiratory
303	1.4	5913	17	T17450	Mutated BRCA1 codi	376	1.4	15223	19	V17553	Respiratory syncyt

377	16	1.4	15223	21	A88743	Respiratory syncyt	c 450	15	1.3	366	21	A42828	Human secreted exp
c 378	16	1.4	16484	20	X13095	Enterococcus faeca	c 451	15	1.3	370	16	T23236	Human gene signatu
c 379	16	1.4	21721	20	X83427	Human lipolysis st	c 452	15	1.3	377	14	O60275	Human brain expres
c 380	16	1.4	22976	20	X83426	Genomic region con	c 453	15	1.3	385	21	C08066	Human secreted pro
c 381	16	1.4	23187	21	A50273	Human lipolysis st	c 454	15	1.3	388	21	C29145	Human secreted pro
c 382	16	1.4	24025	17	T17455	Mutated BRCA1 geno	c 455	15	1.3	389	21	A42471	Human secreted exp
c 383	16	1.4	24025	17	T17515	Mutated BRCA1 geno	c 456	15	1.3	405	20	V90398	EST clone DJ323
c 384	16	1.4	24025	17	T32612	BRCA1, human breas	c 457	15	1.3	428	21	C22866	Human secreted pro
c 385	16	1.4	24026	17	T17512	Mutated BRCA1 geno	c 458	15	1.3	429	21	C02045	Human secreted pro
c 386	16	1.4	24026	17	T17513	Mutated BRCA1 geno	c 459	15	1.3	440	9	N81321	Sequence (I) encod
c 387	16	1.4	24026	17	T17514	Mutated BRCA1 geno	c 460	15	1.3	440	9	N81321	Sequence (I) encod
c 388	16	1.4	24026	17	T17516	Mutated BRCA1 geno	c 461	15	1.3	441	21	C95178	Cat flea head and
c 389	16	1.4	24026	17	T17517	Mutated BRCA1 geno	c 462	15	1.3	441	21	C74686	Human ORFX ORF241
c 390	16	1.4	24026	17	T17518	Mutated BRCA1 geno	c 463	15	1.3	446	20	X13701	Enterococcus faeca
c 391	16	1.4	24026	17	T17519	Mutated BRCA1 geno	c 464	15	1.3	446	20	X13701	EST clone HB443
c 392	16	1.4	24026	17	T17521	Mutated BRCA1 geno	c 465	15	1.3	461	21	C28311	Human secreted pro
c 393	16	1.4	24026	17	T17522	Mutated BRCA1 geno	c 466	15	1.3	461	21	C28311	Polynucleotide seq
c 394	16	1.4	24026	17	T17523	Mutated BRCA1 geno	c 467	15	1.3	462	21	A89636	Exo50 nucleotide s
c 395	16	1.4	24026	17	T17524	Mutated BRCA1 geno	c 468	15	1.3	472	18	T75299	Nucleotide sequenc
c 396	16	1.4	24026	17	T17526	Mutated BRCA1 geno	c 469	15	1.3	472	19	V59980	Nucleic acid encod
c 397	16	1.4	24026	17	T17527	Mutated BRCA1 geno	c 470	15	1.3	486	20	X84692	Human metastatic m
c 398	16	1.4	24026	17	T17528	Mutated BRCA1 geno	c 471	15	1.3	498	21	C93881	Cat flea hindgut a
c 399	16	1.4	24026	17	T17529	Mutated BRCA1 geno	c 472	15	1.3	504	22	F26576	DNA encoding human
c 400	16	1.4	24026	17	T17530	BRCA1, genomic sequ	c 473	15	1.3	533	21	C34677	Arabidopsis thalia
c 401	16	1.4	24026	17	T18325	BRCA1, human breas	c 474	15	1.3	537	21	C34966	Arabidopsis thalia
c 402	16	1.4	24029	17	T17520	Mutated BRCA1 geno	c 475	15	1.3	547	21	T87243	Partial cDNA clone
c 403	16	1.4	24031	17	T17525	Mutated BRCA1 geno	c 476	15	1.3	549	21	C94798	Cat flea hindgut a
c 404	16	1.4	32768	20	X13065	Enterococcus faeca	c 477	15	1.3	550	21	T87242	Partial cDNA clone
c 405	16	1.4	40352	19	V02032	Enterococcus faeca	c 478	15	1.3	551	21	A69773	Human ovarian carc
c 406	16	1.4	78845	21	A81463	MAGE-B cluster DNA	c 479	15	1.3	563	21	C59522	Human secreted pro
c 407	16	1.4	83390	21	F22283	N. meningitidis pa	c 480	15	1.3	563	21	C53809	Arabidopsis thalia
c 408	16	1.4	87350	18	X83003	BAC containing rep	c 481	15	1.3	575	21	T80469	Human colon cancer
c 409	16	1.4	90336	21	F22289	Human WRN genomic	c 482	15	1.3	580	16	T87966	Macaque platelet-a
c 410	16	1.4	349880	21	F21608	BAC containing rep	c 483	15	1.3	580	18	T87069	Macaque platelet-a
c 411	16	1.4	1038602	20	Z01425	Neisseria meningit	c 484	15	1.3	580	18	T80583	Macaque partial pl
c 412	16	1.4	1082138	21	F22305	Complete genome se	c 485	15	1.3	580	19	T96131	Macaque partial pl
c 413	16	1.4	1664976	19	V21209	Arabidopsis thalia	c 486	15	1.3	580	19	T96131	Monkey plasma plat
c 414	16	1.4	1830121	17	T22063	Methanococcus jann	c 487	15	1.3	580	20	X08481	Internal PAF-AH fr
c 415	15	1.3	15	16	T52292	Haemophilus influe	c 488	15	1.3	580	20	V08555	Macaque PAF-AH cod
c 416	15	1.3	15	16	T52522	Mouse ICAM hammerh	c 489	15	1.3	580	21	A59597	DNA encoding plat
c 417	15	1.3	29	21	A57912	Streptococcus pyog	c 490	15	1.3	580	21	A10882	Macaque PAF-AH nuc
c 418	15	1.3	45	16	Q91368	PCR primer for rat	c 491	15	1.3	580	21	A10882	Macaque PAF-AH cdn
c 419	15	1.3	63	16	Q91365	PCR primer for Al.	c 492	15	1.3	580	22	C89075	Platelet-activation
c 420	15	1.3	109	21	C26167	Human secreted pro	c 493	15	1.3	611	18	T83935	DNA encoding a Sta
c 421	15	1.3	117	21	A45581	Human secreted exp	c 494	15	1.3	612	21	C35650	Arabidopsis thalia
c 422	15	1.3	150	21	Z93933	Standard plasmid s	c 495	15	1.3	613	21	A16139	DNA encoding a hum
c 423	15	1.3	150	21	Z93934	Calibrator plasmid	c 496	15	1.3	616	20	X61394	Arabidopsis thalia
c 424	15	1.3	154	16	T20264	Human gene signatu	c 497	15	1.3	627	21	C53531	Bacterial antibiot
c 425	15	1.3	171	21	C16940	Human secreted pro	c 498	15	1.3	630	17	T28570	Polynucleotide seq
c 426	15	1.3	207	21	C16207	Human secreted pro	c 499	15	1.3	632	20	X21129	Human secreted pro
c 427	15	1.3	212	20	X37079	Human cdc37 nuclei	c 500	15	1.3	632	21	C09063	CDNA clone #5 enco
c 428	15	1.3	212	20	X36997	Human cdc37 nuclei	c 501	15	1.3	644	21	T87239	CDNA clone #6 enco
c 429	15	1.3	224	20	X37044	Human cdc37 nuclei	c 502	15	1.3	645	21	T87240	CDNA clone #7 enco
c 430	15	1.3	234	21	C24530	Human secreted pro	c 503	15	1.3	645	21	T87241	Human secreted pro
c 431	15	1.3	239	16	T23798	Human gene signatu	c 504	15	1.3	647	20	X37381	Human secreted pro
c 432	15	1.3	246	21	C30992	Human secreted pro	c 505	15	1.3	649	21	F09445	Fusarium venenatum
c 433	15	1.3	267	16	T19590	Human gene signatu	c 506	15	1.3	650	12	Q13196	TGF-Barrier proteol
c 434	15	1.3	283	21	F62722	Human ORFX ORF1827	c 507	15	1.3	652	19	V03031	Meloidogyne incogn
c 435	15	1.3	294	21	C02207	Human secreted pro	c 508	15	1.3	654	21	A98441	Human PTFN/MWAC1 E
c 436	15	1.3	299	21	T80265	Human colon cancer	c 509	15	1.3	667	21	C75549	Human ORFX ORF1104
c 437	15	1.3	300	20	T12845	Human gene express	c 510	15	1.3	678	21	T87238	CDNA clone #4 enco
c 438	15	1.3	300	20	X98446	Human cancer cell	c 511	15	1.3	688	21	F13709	Aspergillus oryzae
c 439	15	1.3	300	20	A00378	Human colon cancer	c 512	15	1.3	692	20	X13578	Enterococcus faeca
c 440	15	1.3	300	21	A01189	Human colon cancer	c 513	15	1.3	693	21	A70258	Plasmodium falcipa
c 441	15	1.3	300	21	A01421	Human colon cancer	c 514	15	1.3	695	21	T87236	CDNA clone #2 enco
c 442	15	1.3	304	21	A43543	Human secreted exp	c 515	15	1.3	696	16	Q91361	T-cell receptor al
c 443	15	1.3	311	20	Z33443	Human prostate can	c 516	15	1.3	697	21	T87235	CDNA clone #1 enco
c 444	15	1.3	314	16	T26138	Human gene signatu	c 517	15	1.3	706	20	Z17551	Human gene express
c 445	15	1.3	342	21	C26442	Human secreted pro	c 518	15	1.3	706	20	Z17552	Human gene express
c 446	15	1.3	348	18	X83284	Breast cancer tumo	c 519	15	1.3	710	20	Z15825	Human gene express
c 447	15	1.3	348	19	V69006	DNA molecule encod	c 520	15	1.3	711	21	C52426	Arabidopsis thalia
c 448	15	1.3	348	21	C80807	Human breast tumou	c 521	15	1.3	723	19	V31246	E. coli J96 pathog
c 449	15	1.3	348	21	C22584	Human secreted pro	c 522	15	1.3	735	21	Z90654	Nostoc Nsp1 restri

523	15	1.3	740	21	F07748	Fusarium venenatum	596	15	1.3	1386	21	C66032	Human lung cancer-
524	15	1.3	741	21	A98452	Orang utan PTEN/MM	597	15	1.3	1386	21	A08585	Human cytoskeleton
525	15	1.3	741	21	C77586	Human ORFX ORF3141	598	15	1.3	1393	14	C38027	CAD cDNA isolated
526	15	1.3	749	21	A01739	Human colon cancer	599	15	1.3	1393	14	X35655	Arabidopsis thalia
527	15	1.3	750	21	C34456	Arabidopsis thalia	600	15	1.3	1407	21	C33115	Arabidopsis thalia
528	15	1.3	763	21	C39218	Arabidopsis thalia	601	15	1.3	1412	18	T84160	DNA encoding a tob
529	15	1.3	767	21	C40686	Arabidopsis thalia	602	15	1.3	1412	21	C36485	Arabidopsis thalia
530	15	1.3	774	21	C37562	Arabidopsis thalia	603	15	1.3	1427	21	F15577	DNA encoding ome k
531	15	1.3	776	21	C37562	Arabidopsis thalia	604	15	1.3	1436	21	C59245	Arabidopsis thalia
532	15	1.3	776	21	C37562	Arabidopsis thalia	605	15	1.3	1439	14	Q50486	Human prostate can
533	15	1.3	782	21	A01907	Human signal pepti	606	15	1.3	1440	17	T28571	Human secreted pro
534	15	1.3	783	20	X39955	prostate cancer as	607	15	1.3	1449	17	T28571	S.cerevisiae ribof
535	15	1.3	799	21	F08376	Fusarium venenatum	608	15	1.3	1451	21	C50666	Bacterial antibiot
536	15	1.3	806	21	C52603	Arabidopsis thalia	609	15	1.3	1451	21	C35262	Arabidopsis thalia
537	15	1.3	827	20	Z17167	Human gene express	610	15	1.3	1451	21	C38898	KREL coding sequen
538	15	1.3	827	21	C48824	Human gene express	611	15	1.3	1465	21	C45984	Arabidopsis thalia
539	15	1.3	843	20	C48824	Human gene express	612	15	1.3	1467	21	C37381	Arabidopsis thalia
540	15	1.3	847	20	X39814	Soybean Drl coding	613	15	1.3	1519	20	Z10674	Arabidopsis thalia
541	15	1.3	851	21	C53810	Gastric cancer ass	614	15	1.3	1520	21	C64786	CAD encoding a hu
542	15	1.3	852	21	A81913	Arabidopsis thalia	615	15	1.3	1520	21	C64786	Lemon acyl transfe
543	15	1.3	864	20	X13704	N. meningitidis pa	616	15	1.3	1534	18	X30755	Streptococcus pneu
544	15	1.3	906	21	A53723	Enterococcus faeca	617	15	1.3	1542	19	C37391	Streptococcus pneu
545	15	1.3	959	22	C89717	Campylobacter jeju	618	15	1.3	1545	21	C49085	Human ikappaB poly
546	15	1.3	988	21	A63737	Human cytoskeleton	619	15	1.3	1548	19	V5840	Encodes I-kappa-B-al
547	15	1.3	999	20	Z17304	DNA encoding a pol	620	15	1.3	1550	13	Q31550	Human I-kappa-B-al
548	15	1.3	1001	21	C57628	Human gene express	621	15	1.3	1550	13	Q31550	Human I-kappa-B-al
549	15	1.3	1001	21	C57629	Arachidonic acid m	622	15	1.3	1551	20	A51230	Human cell regulat
550	15	1.3	1001	21	C57630	Arachidonic acid m	623	15	1.3	1551	20	A51230	Human lung cancer-
551	15	1.3	1001	21	C57661	Arachidonic acid m	624	15	1.3	1551	20	A51230	DNA encoding a S.
552	15	1.3	1001	21	C57661	Arachidonic acid m	625	15	1.3	1551	20	A51230	Human mesoderm ind
553	15	1.3	1019	21	C36901	Optimum primer can	626	15	1.3	1574	20	Z08075	qDNA encoding a hu
554	15	1.3	1024	21	Z97146	Human secreted pro	627	15	1.3	1574	20	Z08075	T668K-2 gene seque
555	15	1.3	1024	21	Z97146	Human secreted pro	628	15	1.3	1574	20	Z08075	Basal endosperm tr
556	15	1.3	1034	18	T73216	Human protein coding	629	15	1.3	1574	20	Z08075	Arabidopsis thalia
557	15	1.3	1034	18	T73216	Glucagon-like pept	630	15	1.3	1615	20	X90969	Sequence encoding
558	15	1.3	1034	18	T73216	Rat prepro-glucago	631	15	1.3	1615	20	X90969	Bacterial antibiot
559	15	1.3	1034	18	T73216	Preproglucagon cod	632	15	1.3	1615	20	X90969	Enterococcus faeca
560	15	1.3	1102	16	Q88812	Clone pBP352. Syn	633	15	1.3	1615	20	X90969	Neisseria gonorrhoe
561	15	1.3	1104	19	X14156	H. pylori GHPO 461	634	15	1.3	1615	20	X90969	Arabidopsis thalia
562	15	1.3	1109	19	V59640	Human secreted pro	635	15	1.3	1615	20	X90969	Human secreted pro
563	15	1.3	1109	19	V59640	Human secreted pro	636	15	1.3	1615	20	X90969	Arabidopsis thalia
564	15	1.3	1111	21	C35479	Arabidopsis thalia	637	15	1.3	1615	20	X90969	Arabidopsis thalia
565	15	1.3	1111	21	C35479	Arabidopsis thalia	638	15	1.3	1615	20	X90969	Arabidopsis thalia
566	15	1.3	1118	21	C52804	Arabidopsis thalia	639	15	1.3	1615	20	X90969	Human cell regulat
567	15	1.3	1127	21	C47425	Arabidopsis thalia	640	15	1.3	1615	20	X90969	Nucleotide sequenc
568	15	1.3	1129	21	C36437	Arabidopsis thalia	641	15	1.3	1615	20	X90969	Shl gene contain
569	15	1.3	1130	21	Z98188	Human signal pepti	642	15	1.3	1615	20	X90969	cDNA encoding a hu
570	15	1.3	1161	9	N81711	Sequence of fowlpo	643	15	1.3	1615	20	X90969	Human endokine alp
571	15	1.3	1163	21	C34345	Arabidopsis thalia	644	15	1.3	1615	20	X90969	T7 bacteriophage 3
572	15	1.3	1165	21	C35434	Arabidopsis thalia	645	15	1.3	1615	20	X90969	Human transmembran
573	15	1.3	1167	21	C50158	Arabidopsis thalia	646	15	1.3	1615	20	X90969	Nucleic acid seque
574	15	1.3	1175	21	C36832	Arabidopsis thalia	647	15	1.3	1615	20	X90969	Nucleotide sequenc
575	15	1.3	1182	20	X58576	Human cell regulat	648	15	1.3	1615	20	X90969	Riboflavin 5 gene
576	15	1.3	1191	16	Q92862	Thermotable alkali	649	15	1.3	1615	20	X90969	Human pancreatic c
577	15	1.3	1209	21	C47056	Arabidopsis thalia	650	15	1.3	1615	20	X90969	Arabidopsis thalia
578	15	1.3	1251	21	A05866	Clathrin-adaptor p	651	15	1.3	1615	20	X90969	Arabidopsis thalia
579	15	1.3	1262	18	V74767	Group B Streptococ	652	15	1.3	1615	20	X90969	Human cell regulat
580	15	1.3	1294	21	C38600	Staphylococcus aur	653	15	1.3	1615	20	X90969	Sequence encoding
581	15	1.3	1302	21	A65975	Arabidopsis thalia	654	15	1.3	1615	20	X90969	Essential Staphylo
582	15	1.3	1308	18	V74387	Staphylococcus aur	655	15	1.3	1615	20	X90969	Human PRO364 ligan
583	15	1.3	1311	21	C9126	E. coli proliferat	656	15	1.3	1615	20	X90969	cDNA sequence for
584	15	1.3	1320	20	X39657	Staphylococcus aur	657	15	1.3	1615	20	X90969	Human PRO175 cDNA
585	15	1.3	1336	19	V04079	Arabidopsis thalia	658	15	1.3	1615	20	X90969	Human angiologenesis
586	15	1.3	1336	19	V04079	Arabidopsis thalia	659	15	1.3	1615	20	X90969	Human cancer assoc
587	15	1.3	1337	21	C42006	Renal cancer assoc	660	15	1.3	1615	20	X90969	DNA encoding a pol
588	15	1.3	1337	21	C49513	Aplysia CAMP-respo	661	15	1.3	1615	20	X90969	Human secreted pro
589	15	1.3	1345	18	T59939	Arabidopsis thalia	662	15	1.3	1615	20	X90969	Human secreted pro
590	15	1.3	1347	20	X58574	Phage resistance g	663	15	1.3	1615	20	X90969	Human p40 coding s
591	15	1.3	1351	19	X14005	Human cell regulat	664	15	1.3	1615	20	X90969	Human lung cancer-
592	15	1.3	1352	17	C30570	Eukaryotic cell gr	665	15	1.3	1615	20	X90969	Eucalyptus grandis
593	15	1.3	1353	21	C34381	Arabidopsis thalia	666	15	1.3	1615	20	X90969	Lymphocyte chemoat
594	15	1.3	1355	21	C46931	Arabidopsis thalia	667	15	1.3	1615	20	X90969	
595	15	1.3	1372	21	C34826	Arabidopsis thalia	668	15	1.3	1615	20	X90969	
			1377	21	C43415	Arabidopsis thalia							

669	1.3	2150	18	T49000	Lymphocyte chemoat	c 742	15	1.3	3150	14	Q51428	Human FACC cDNA cl
670	1.3	2151	17	T33895	Lymphocyte chemoat	743	15	1.3	3175	21	F21348	Human low adenosin
671	1.3	2171	9	N81668	Sequence of part o	744	15	1.3	3175	21	A35226	Human adenosine re
c 672	1.3	2179	19	T98571	DNA encoding a S.	745	15	1.3	3200	21	A61503	DNA encoding the C
c 673	1.3	2206	15	Q70754	Beta tubulin gene	746	15	1.3	3213	16	Q79870	Lactobacillus curv
c 674	1.3	2229	21	A16696	Human secreted pro	747	15	1.3	3213	16	Q79869	Lactobacillus curv
c 675	1.3	2232	21	C43218	Arabidopsis thalia	748	15	1.3	3265	20	V95563	Maize lysine keto
c 676	1.3	2232	21	Z52444	HTM clone 2474110	749	15	1.3	3359	19	V59078	Soybean seed coat
c 677	1.3	2257	20	X76839	T66k-8 gene sequen	750	15	1.3	3417	21	C43064	Arabidopsis thalia
678	1.3	2270	20	Z25771	Human p51 encoding	751	15	1.3	3435	16	Q79746	M. catarrhalis 422
679	1.3	2270	21	C66028	Human lung cancer-	752	15	1.3	3438	16	T95247	Acremonium chrysog
c 680	1.3	2286	21	C42714	Arabidopsis thalia	753	15	1.3	3445	14	Q82230	A. chrysogenum beta
c 681	1.3	2302	20	X93863	Human secreted pro	754	15	1.3	3445	14	Q55405	A. chrysogenum beta
c 682	1.3	2311	21	F18225	Lung cancer associ	755	15	1.3	3445	14	Q55406	Nucleic acid encod
c 683	1.3	2312	21	C80535	Human secreted pro	756	15	1.3	3455	20	V81385	Arabidopsis thalia
c 684	1.3	2325	19	Z96293	S. pneumoniae deri	757	15	1.3	3455	20	C46960	Soybean DNA-bindin
685	1.3	2334	21	C39820	Arabidopsis thalia	758	15	1.3	3528	18	T94547	Human low adenosin
c 686	1.3	2344	14	Q51427	Human FACC cDNA cl	759	15	1.3	3568	21	F21349	Human adenosine re
687	1.3	2362	19	V16530	DNA encoding a Bac	760	15	1.3	3568	21	A35227	M. catarrhalis Q8
688	1.3	2362	20	X83891	Bacillus thuringie	761	15	1.3	3660	18	T95249	Phage resistance g
689	1.3	2367	19	V16525	DNA encoding a Bac	762	15	1.3	3704	18	T59938	Deltex cDNA. Homo
690	1.3	2367	20	X83886	Bacillus thuringie	763	15	1.3	3771	16	Q92980	Drosophila deltex
691	1.3	2370	19	V16519	Bacillus thuringie	764	15	1.3	3771	18	T68924	Staphylococcus aur
692	1.3	2370	19	X83880	DNA encoding a Bac	765	15	1.3	3771	18	T68924	Human polynucleoti
693	1.3	2370	20	X83889	Bacillus thuringie	766	15	1.3	3813	22	C91329	Nucleotide sequenc
694	1.3	2374	20	X30288	Bacillus thuringie	767	15	1.3	3813	22	A59329	Human STAT1-cDNA.
695	1.3	2375	19	V30289	Bacillus thuringie	768	15	1.3	3943	17	T31276	Human STAT1-alpha
696	1.3	2375	19	V30290	Bacillus thuringie	769	15	1.3	3943	17	Q49165	Human STAT1-cDNA.
697	1.3	2375	19	V16528	DNA encoding a Bac	770	15	1.3	3977	14	Q49165	91 KD ISGF-3alpha
698	1.3	2375	19	V16528	DNA encoding a Bac	771	15	1.3	4002	20	X85823	cDNA encoding mult
699	1.3	2375	20	X83881	Bacillus thuringie	772	15	1.3	4003	21	F21052	Human low adenosin
700	1.3	2375	21	Z58782	DNA encoding a tox	773	15	1.3	4003	21	A34930	Human adenosine re
701	1.3	2376	20	X83884	DNA encoding a Bac	774	15	1.3	4003	22	A89227	Human signal trans
702	1.3	2377	10	N90726	Bacillus thuringie	775	15	1.3	4328	18	V74590	Staphylococcus aur
703	1.3	2377	10	N90726	DNA sequence of ex	776	15	1.3	4328	18	V74590	Pyruvate:flavodoxi
704	1.3	2377	15	Q68538	Extracellular neut	777	15	1.3	4387	21	C77052	Human ORFX ORF2607
705	1.3	2378	17	T13942	B. thuringiensis v	778	15	1.3	4465	17	T42135	12C-2 gene encodin
706	1.3	2378	18	T74003	B. cereus VIP3A(a)	779	15	1.3	4465	17	T42135	DNA encoding a cal
707	1.3	2378	19	V16176	Native DNA sequenc	780	15	1.3	4488	14	Q51426	Human FACC cDNA cl
c 708	1.3	2394	21	C77076	Human ORFX ORF2631	781	15	1.3	4549	21	Z38981	dhat and dhaB(1,2,
709	1.3	2463	21	C78193	Human cancer assoc	782	15	1.3	4549	21	Z38981	Fanconi anaemia co
710	1.3	2481	20	X87652	Mouse interleukin-	783	15	1.3	4567	21	C62127	DNA encoding a cal
c 711	1.3	2522	19	V16211	cDNA encoding mous	784	15	1.3	4611	21	C62127	Human lung tumor a
712	1.3	2612	17	T13943	B. thuringiensis v	785	15	1.3	4611	21	C62127	Human lung cancer-
713	1.3	2612	18	T74005	B. cereus VIP3A(b)	786	15	1.3	4655	21	V30461	Soybean seed coat
714	1.3	2612	19	V68063	Bacillus thuringie	787	15	1.3	4700	19	V30461	Nucleotide sequenc
715	1.3	2612	19	V68063	Native DNA sequenc	788	15	1.3	4700	20	Z29902	Human ORFX ORF2063
716	1.3	2671	17	T33648	Ras p21 interactin	789	15	1.3	4711	21	C76508	Human KET cDNA. H
717	1.3	2671	20	Z09608	Human RGL cDNA. H	790	15	1.3	4846	21	F243913	Human lung cancer-
718	1.3	2715	21	A70255	Plasmodium falcipa	791	15	1.3	4846	21	C66031	12C-1 gene encodin
719	1.3	2723	8	N70197	Sequence of BAR 1	792	15	1.3	4948	17	T42134	P. fluorescens DNA
c 720	1.3	2738	21	C47165	Arabidopsis thalia	793	15	1.3	5057	20	Z23757	Deltex locus compo
721	1.3	2750	10	N90722	Nucleotide sequenc	794	15	1.3	5057	20	Z23757	Arabidopsis thalia
722	1.3	2750	12	Q13195	Barrier protein BA	795	15	1.3	5259	21	C50213	Rat thyrotropin re
723	1.3	2750	13	Q20266	BAR1 barrier prote	796	15	1.3	5270	13	Q25387	Plasmid pGM707 con
724	1.3	2769	22	F44697	Novel protein kina	797	15	1.3	5270	13	Q25387	Plasmid pGM708 con
c 725	1.3	2774	21	C77613	Human cancer assoc	798	15	1.3	5356	21	A51631	Plasmid vector pSE
726	1.3	2816	20	Z25770	Human p51 encoding	799	15	1.3	5356	21	A51631	Staphylococcus aur
727	1.3	2816	21	C66029	Human lung cancer-	800	15	1.3	5763	18	V74816	Staphylococcus aur
728	1.3	2820	21	C66029	Human lung cancer-	801	15	1.3	5804	18	V74816	Hydra head activat
729	1.3	2889	21	A70145	Plasmodium falcipa	802	15	1.3	5839	19	V46335	Mouse A-myb genom
730	1.3	2893	22	C66890	Human EXMAD-1 codi	803	15	1.3	5889	20	V84328	Plasmid pGM769 con
c 731	1.3	2894	17	T10423	H. influenzae SB33	804	15	1.3	5889	20	V84328	pGM807 with udp-de
c 732	1.3	2967	21	X82625	Human secreted pro	805	15	1.3	5889	20	V84328	Enterococcus faeca
c 733	1.3	2981	21	X82625	Human IGA nephrop	806	15	1.3	6117	21	A51644	Plasmid pGM716 con
734	1.3	3045	18	T73866	Cotton fibre promo	807	15	1.3	6117	21	A51644	IAMS expression ve
735	1.3	3101	18	T99905	H. pylori GHPo 111	808	15	1.3	6301	21	A51635	Plasmid pGM771 con
736	1.3	3106	18	T99905	Enterohaemorrhagic	809	15	1.3	6343	21	A51638	Staphylococcus aur
737	1.3	3106	18	T99905	Enterohaemorrhagic	810	15	1.3	6343	21	A51638	Tomato immunity 2
c 738	1.3	3123	21	A65930	E. coli proliferat	811	15	1.3	6444	18	V74544	Human N-proteinase
739	1.3	3131	18	T99904	EHEC eae gene asso	812	15	1.3	6691	21	Z57204	Human N-proteinase
740	1.3	3131	18	T99904	Enterohaemorrhagic	813	15	1.3	6691	21	Z57204	T-DNA containing t
741	1.3	3131	18	V02322	Enterohaemorrhagic	814	15	1.3	6743	20	X32515	Cinnamyl alcohol d

ALIGNMENTS

DR WPI; 1999-215066/18.

QY	1	gcatacgtggtcgacatagctctcaagacagataactctggaaaaaacatcacagaacgtgctt	60
DG	1	gcatacgtggtcgacatagctctcaagacagataactctggaaaaaacatcacagaacgtgctt	60
QY	61	gacaagtcttgagatagaaaaaaacatcgtttaaaacctaaagataagtagtgaacttaagtttaatt	120
DG	61	gacaagtcttgagatagaaaaaaacatcgtttaaaacctaaagataagtagtgaacttaagtttaatt	120
QY	121	aataaagcgtttaaaaagatacataatgatgatgttaactgaagagtgtagtgaatgctaagaag	180
DG	121	aataaagcgtttaaaaagatacataatgatgatgttaactgaagagtgtagtgaatgctaagaag	180
QY	181	aaactacgttgatccgcgcgtgactaggggtacataataatgcccgcaaaagagcaaaagaa	240
DG	181	aaactacgttgatccgcgcgtgactaggggtacataataatgcccgcaaaagagcaaaagaa	240
QY	241	gctcttgacaagtatgagctagagaaaaaccatgacttaaaaaactgaagaatgaagggttcaaaa	300
DG	241	gctcttgacaagtatgagctagagaaaaaccatgaacttaaaaaactgaagaatgaagggttcaaaa	300
QY	301	actgagaataagagggtttaaaactgagaatgaagggtttaaaaaactgagaatgaagggttca	360
DG	301	actgagaaatgaagggtttaaaaaactgagaatgaagggtttaaaaaactgagaatgaagggttca	360
QY	361	aaaactgaggtcgacagagtggttcttcaggggggacggtagaaaaaccgggacaaagcacga	420
DG	361	aaaactgaggtcgacagagtggttcttcaggggggacggtagaaaaaccgggacaaagcacga	420
QY	421	gaactcttaacaagtatacagtagaagaacttatgtttacaagctaatatgacaaagtta	480
DG	421	gaactcttaacaagtatgacgtagagaactctatgtttacaagctaatatgacaaagtta	480
QY	481	ccatggagagtgcggttatactactaggcatacgcgcagaagaataagcttaaaaaaaattattgac	540
DG	481	ccatggagagtgcggttatactactaggcatacgcgcagaagaataagcttaaaaaaaattattgac	540
QY	541	gatcttgacgcaaaaagaacatagaattacaaacaaagatatgagaagttatctctgcagaac	600
DG	541	gatcttgacgcaaaaagaacatagaattacaaacaaagatatgagaagttatctctgcagaac	600
QY	601	ggtagtgttaactcctagggaagttaaagaagatctctgcagcaaaacaaatccccgcaatacaa	660
DG	601	ggtagtgttaactcctagggaagttaaagaagatctctgcagcaaaacaaatccccgcaatacaa	660
QY	661	aatacgtttacgtcacgaaaaaacaggacttaaaagcgagatgagagaatgcaattggaa	720
DG	661	aatacgtttacgtcacgaaaaaacaggacttaaaagcgagatgagagaatgcaattggaa	720

```
QY 721 gttcaggaagagattttaagagagctggtaccttgttagatcaggttacacaattat 780
Db 721 gttcaggaagagattttaagagagctggtaccttgttagatcaggttacacaattat 780
QY 781 actaaacataatagtaattaccaacaataataatgacaaagctggcagacttgacctgaga 840
Db 781 actaaacataatagtaattaccaacaataataatgacaaagctggcagacttgacctgaga 840
QY 841 caaaagctgaatatctaaagcccttaattgattggctgagagctgttacaagagtta 900
Db 841 caaaagctgaatatctaaagcccttaattgattggctgagagctgttacaagagtta 900
QY 901 aatatcgatgcgaactaggtctcagacagatactctggaaaaagtacaaagaacgtgct 960
Db 901 aatatcgatgcgaactaggtctcagacagatactctggaaaaagtacaaagaacgtgct 960
QY 961 gacaagtttgatgagaaaacataactgttaaaacttagaataagtgacttaagttaat 1020
Db 961 gacaagtttgatgagaaaacataactgttaaaacttagaataagtgacttaagttaat 1020
QY 1021 aataaagcgttaaaagataataatgatgagtttaactgaagctgaggttaagtctaaagag 1080
Db 1021 aataaagcgttaaaagataataatgatgagtttaactgaagctgaggttaagtctaaagag 1080
QY 1081 aaactacgtataaataatcactatctgaaaaagctagtaaaattcagaattagag 1140
Db 1081 aaactacgtataaataatcactatctgaaaaagctagtaaaattcagaattagag 1140
QY 1141 gcagctaaagtaaaagcctt 1158
Db 1141 gcagctaaagtaaaagcctt 1158
```

RESULT 2

A57894
ID A57894 standard; DNA; 741 BP.

XX AC A57894;

DT 10-OCT-2000 (first entry)

DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:3.

XX KW Multivalent hybrid M protein; group A streptococcus; serotype; immunogenic; sero-specific antibody; streptococcal infection; cross reactivity; vaccine; acute rheumatic fever; ARF; rheumatic heart disease; streptococcal pharyngitis; strep throat; pneumonia; ds.

XX OS Streptococcus pyogenes.

XX OS Synthetic.

XX PN US6063386-A.

XX PD 16-MAY-2000.

XX PF 15-SEP-1997; 97US-0937271.

XX PR 16-SEP-1992; 92US-0945954.

XX PA (UYTE-) UNIV TENNESSEE RES CORP.

XX PI Lederer JW, Dale JB;

XX DR WPI; 2000-364475/31.

XX DR P-PSDB; B03114.

XX PT New immunogenic recombinant hybrid M protein comprising amino-terminal peptide fragments of streptococcal M protein useful as vaccine against rheumatic fever and infections leading to rheumatic fever

XX PS Disclosure; Fig 4A-B; 62pp; English.

XX

CC The invention relates to multivalent immunogenic hybrid group A streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19. The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a multivalent hybrid M protein; and a method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences A57893-A57902 represent DNAs encoding multivalent hybrid Streptococcus pyogenes M proteins generated in the disclosure of the invention.

SQ Sequence 741 BP; 311 A; 106 C; 164 G; 160 T; 0 other;

Query Match

Best Local Similarity 34.7%; Score 402; DB 21; Length 741;

Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 190 ggatccgcgctgactagggttacataataatgaccccaagacaaagaagctcttgac 249
Db 340 ggatccgcgctgactagggttacataataatgaccccaagacaaagaagctcttgac 399
QY 250 aagtatgactagaaaaccatgacttaaaactaagaatgaaggggttaaaactgagaat 309
Db 400 aagtatgactagaaaaccatgacttaaaactaagaatgaaggggttaaaactgagaat 459
QY 310 gaaggggttaaaactgagaatgaaggggttaaaactaagaatgaaggggttaaaactgaga 369
Db 460 gaaggggttaaaactgagaatgaaggggttaaaactaagaatgaaggggttaaaactgaga 519
QY 370 gtcgacagagtgttctctaggggacggtagaaaacccggacaaagcacgagaactctt 429
Db 520 gtcgacagagtgttctctaggggacggtagaaaacccggacaaagcacgagaactctt 579
QY 430 aacaagtatgactagagaactctatgttacaagctaatataatgacaagtatcatggaga 489
Db 580 aacaagtatgactagagaactctatgttacaagctaatataatgacaagtatcatggaga 639
QY 490 gtgcgttactaggcatacgccagaagaataagctaaaaaaattattgacgactcttgac 549
Db 640 gtgcgttactaggcatacgccagaagaataagctaaaaaaattattgacgactcttgac 699
QY 550 gcaaaagaacatgaattacaacaacagaatgagaagttatct 591
Db 700 gcaaaagaacatgaattacaacaacagaatgagaagttatct 741
```

RESULT# 3

Q45215

ID Q45215 standard; DNA; 741 BP.

XX AC Q45215;

XX AC Q45215;

DT 02-NOV-1994 (first entry)

XX DE Recombinant M24-M5-M6-M19 gene.

XX

|||||
580 aacagtagcgtagagactctatgttacagctaatgaacacattaccatggaga 639
|||||
490 gtgcgttatactagcgcagcagagataagcctaaacaaattattgacgatcttgac 549
|||||
640 gtgcgttatactagcgcagcagagataagcctaaacaaattattgacgatcttgac 699
|||||
550 gcaaaagaacatgaattacacacacagaatgagaagtattct 591
|||||
700 gcaaaagaacatgaattacacacacagaatgagaagtattct 741
|||||
RESULT 4
Q45214
ID Q45214 standard; DNA; 861 BP.
XX
AC Q45214;
XX
DT 02-NOV-1994 (first entry)
XX
DE Recombinant M24-M5-M6 gene.
XX
KW Primer: PCR: amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_difference 781..783
FT /*tag= a
FT /codon= seq:gga, aa:Arg
XX
PN WO9406421-A.
XX
PD 31-MAR-1994.
XX
PF 15-SEP-1993; 93WO-US08703.
XX
PR 16-SEP-1992; 92US-0945954.
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX Dale JB, Lederer JW;
XX PI
XX WPI: 1994-118122/14.
XX P-PSDB; R50992.
XX
PT New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes
XX
PS Disclosure: Fig 1; 67pp; English.
XX
CC The sequences given in Q45214-23 encode hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include

KW Primer: PCR: amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal; ss.
XX
OS Synthetic.
XX
PN WO9406421-A.
XX
PD 31-MAR-1994.
XX
PF 15-SEP-1993; 93WO-US08703.
XX
PR 16-SEP-1992; 92US-0945954.
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
DE Dale JB, Lederer JW;
XX
KW WPI: 1994-118122/14.
KW P-PSDB; R50993.
XX
CC New immunogenic hybrid proteins derives from streptococcal M proteins
CC PT - induces opsonic antibodies, for protective immunisation against
CC PT against multiple group A streptococci serotypes
XX
PS Disclosure: Fig 4; 67pp; English.
XX
CC The sequences given in Q45214-23 encode hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.
XX
SQ Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;

Query Match 30.38; Score 351; DB 15; Length 741;
Best Local Similarity 99.88; Pred. No. 9.4e-165;
Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 190 gqatccgcgtgactaggggtacataataatgaccgcgcaagagcaaaagagctcttgac 249
|||||
Db 340 gqatccgcgtgactaggggtacataataatgaccgcgcaagagcaaaagagctcttgac 399
|||||
QY 250 aagtagagctagaaacacagcttaaaactaaagaggttaaaactgagaat 309
|||||
Db 400 aagtagagctagaaacacagcttaaaactaaagaggttaaaactgagaat 459
|||||
QY 310 gaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactgag 369
|||||
Db 460 gaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactgag 519
|||||
QY 370 gtgcacagagtggttcttagggcggttagaaacccgagcaaacgacgagactctt 429
|||||
Db 520 gtgcacagagtggttcttagggcggttagaaacccgagcaaacgagcaaacgactctt 579
|||||
QY 430 aacagtagcgtagagactctatgttacagctaatgaacacattaccatggaga 489

CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX
 SQ Sequence 861 BP; 331 A; 147 C; 199 G; 184 T; 0 other;

Query Match 20.7%; Score 240; DB 15; Length 861;
 Best Local Similarity 100.0%; Pred. No. 1.3e-109;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 910 gtcgcgactaggtctcagacagatctctgaaaaagctacaagaacgctgctgacaagttt 969
 Db 4 gtcgcgactaggtctcagacagatctctgaaaaagctacaagaacgctgctgacaagttt 63
 QY 970 gagatagaaaaacaatacgttaaaacttaagaatagtgacttaagtttttaataataaagcg 1029
 Db 64 gagatagaaaaacaatacgttaaaacttaagaatagtgacttaagtttttaataataaagcg 123
 QY 1030 ttaaaagatcataatgatgatgtaactgaagattgagtaactgaagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgatgtaactgaagattgagtaactgaagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 1149
 Db 184 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 243

RESULT 5
 A57893
 ID A57893 standard; DNA; 861 BP.
 XX
 AC A57893;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE S. Pyogenes hybrid M protein (M24-M5-M6) DNA, SEQ ID NO:1.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia; ds.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX
 PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 15-SEP-1997; 97US-0937271.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Lederer JW, Dale JB;
 XX
 DR WPI; 2000-364475/31.
 DR P-PSDB; B03113.
 XX
 PT New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever
 XX
 PS Disclosure; Fig 1A-B; 62pp; English.
 XX
 CC The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences A57893-A57902
 CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
 CC proteins generated in the disclosure of the invention.

XX Sequence 861 BP; 331 A; 147 C; 199 G; 184 T; 0 other;
 SQ

Query Match 20.7%; Score 240; DB 21; Length 861;
 Best Local Similarity 100.0%; Pred. No. 1.3e-109;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 910 gtcgcgactaggtctcagacagatctctgaaaaagctacaagaacgctgctgacaagttt 969
 Db 4 gtcgcgactaggtctcagacagatctctgaaaaagctacaagaacgctgctgacaagttt 63
 QY 970 gagatagaaaaacaatacgttaaaacttaagaatagtgacttaagtttttaataataaagcg 1029
 Db 64 gagatagaaaaacaatacgttaaaacttaagaatagtgacttaagtttttaataataaagcg 123
 QY 1030 ttaaaagatcataatgatgatgtaactgaagattgagtaactgaagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgatgtaactgaagattgagtaactgaagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 1149
 Db 184 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 243

RESULT 6
 Q45220
 ID Q45220 standard; DNA; 522 BP.
 XX
 AC Q45220;
 XX
 DT 02-NOV-1994 (first entry)
 XX
 DE Recombinant M24-M5 divalent hybrid gene.
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX
 OS Synthetic.
 XX
 PN WO9406421-A.
 XX
 PD 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.

PI Dale JB, Lederer JW;
 XX WPI; 1994-118122/14.
 DR P-PSDB; R50998.
 XX
 DR New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX
 XX Disclosure; Fig 10; 67pp; English.
 PS
 XX The sequences given in Q45214-23 encode hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC PK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX
 SQ Sequence 522 BP; 226 A; 68 C; 116 G; 112 T; 0 other;

Query Match 18.7%; Score 217; DB 15; Length 522;
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 910 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 969
 Db 4 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 63
 QY 970 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttttaaataaagcg 1029
 Db 64 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttttaaataaagcg 123
 QY 1030 ttaaaagatcataatgatgatgttaactgaagagttgagtaactgaagagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgatgttaactgaagagttgagtaactgaagagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaagcgtagtaaaa 1126
 Db 184 aaaaatgataaatacactatctgaaaagcgtagtaaaa 220

RESULT 7
 A57899
 ID A57899 standard; DNA; 522 BP.
 XX
 AC A57899;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE S. pyogenes hybrid M protein (M24-M5) DNA, SEQ ID NO:14.
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia; ds.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX

PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 15-SEP-1997; 97US-0937271.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Lederer JW, Dale JB;
 XX
 DR WPI; 2000-364475/31.
 DR P-PSDB; B03120.
 XX
 XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever;
 XX
 XX Disclosure; Fig 10A-B; 62pp; English.
 PS
 XX The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences A57893-A57902
 CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
 CC proteins generated in the disclosure of the invention.
 XX
 SQ Sequence 522 BP; 226 A; 68 C; 115 G; 113 T; 0 other;

Query Match 18.7%; Score 217; DB 21; Length 522;
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 910 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 969
 Db 4 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 63
 QY 970 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttttaaataaagcg 1029
 Db 64 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttttaaataaagcg 123
 QY 1030 ttaaaagatcataatgatgatgttaactgaagagttgagtaactgaagagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgatgttaactgaagagttgagtaactgaagagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaagcgtagtaaaa 1126
 Db 184 aaaaatgataaatacactatctgaaaagcgtagtaaaa 220

ID Q45219 standard; DNA: 741 BP.
 AC Q45219;
 XX
 DT 02-NOV-1994 (first entry)
 XX
 DE Recombinant M19-M6-M5-M24 tetraivalent hybrid gene.
 XX
 KW Primer: PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; pkk223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX
 OS Synthetic.
 XX
 PN WO9406421-A.
 XX
 PD 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 DR WPI: 1994-118122/14.
 DR P-PSDB; R50997.
 XX
 PT New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX
 PS Disclosure; Fig 9; 67pp; English.
 XX
 CC The sequences given in Q45214-23 encode hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC pKK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX
 SQ Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;
 Query Match 18.7%; Score 217; DB 15; Length 741;
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacgatactctggaaaaagtacaaagcgtctgacaagttt 969
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 406 gtcgcgactaggtctcagacgatactctggaaaaagtacaaagcgtctgacaagttt 465
 QY 970 gagatagaaacaatacgtttaaactaaagtaagtacttaagtttttaataaagcg 1029
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 466 gagatagaaacaatacgtttaaactaaagtaagtacttaagtttttaataaagcg 525
 QY 1030 ttaaagaatcataatgatgagtttaactgaagagttgagtaactgaagaaactacgt 1089

Db 526 ttaaagaatcataatgatgagtttaactgaagagttgagtaactgaagaaactacgt 585
 QY 1090 aaaaatgataaactactatctgaaaaagcgtagtaaaa 1126
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 586 aaaaatgataaactactatctgaaaaagcgtagtaaaa 622
 RESULT 9
 A57898
 ID A57898 standard; DNA: 741 BP.
 XX
 AC A57898;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:12.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia; ds.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX
 PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 15-SEP-1997; 97US-0937271.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Lederer JW, Dale JB;
 DR WPI: 2000-364475/31.
 DR P-PSDB; B03119.
 XX
 PT New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever
 XX
 PS Disclosure; Fig 9A-B; 62pp; English.
 XX
 CC The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences A57893-A57902
 CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
 CC proteins generated in the disclosure of the invention.

CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX
 SQ Sequence 822 BP; 332 A; 136 C; 182 G; 172 T; 0 other;

Query Match 18.7%; Score 217; DB 21; Length 741;
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaagaaagtagtgaactgtgctgacaagttt 969
 Db 4' gtcgcgactaggtctcagacagatactctggaagaaagtagtgaactgtgctgacaagttt 63
 QY 970 gagatagaaaaacaatacgtttaaaacttaagaataagtagtgaactgtttaaataaagcg 1029
 Db 64 gagatagaaaaacaatacgtttaaaacttaagaataagtagtgaactgtttaaataaagcg 123
 QY 1030 ttaaaagatcataatgatgatgagtgtaactgaagagtgtagtgaactgaagagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgatgagtgtaactgaagagtgtagtgaactgaagagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaagcgtagtaaaa 1126
 Db 184 aaaaatgataaatacactatctgaaaagcgtagtaaaa 220

RESULT 11
 AS7895
 ID AS7895 standard; DNA; 822 BP.
 XX
 AC AS7895;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:5.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia; ds.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 15-SEP-1997; 97US-0937271.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Lederer JW, Dale JB;
 XX
 DR WPI; 2000-364475/31.
 DR P-PSDB; B03115.
 XX
 PT New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever -
 XX

XX
 SQ Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;

Query Match 18.7%; Score 217; DB 21; Length 741;
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaagaaagtagtgaactgtgctgacaagttt 969
 Db 406 gtcgcgactaggtctcagacagatactctggaagaaagtagtgaactgtgctgacaagttt 465
 QY 970 gagatagaaaaacaatacgtttaaaacttaagaataagtagtgaactgtttaaataaagcg 1029
 Db 466 gagatagaaaaacaatacgtttaaaacttaagaataagtagtgaactgtttaaataaagcg 525
 QY 1030 ttaaaagatcataatgatgatgagtgtaactgaagagtgtagtgaactgaagagaaactacgt 1089
 Db 526 ttaaaagatcataatgatgatgagtgtaactgaagagtgtagtgaactgaagagaaactacgt 585
 QY 1090 aaaaatgataaatacactatctgaaaagcgtagtaaaa 1126
 Db 586 aaaaatgataaatacactatctgaaaagcgtagtaaaa 622

RESULT 10
 Q45216
 ID Q45216 standard; DNA; 822 BP.
 XX
 AC Q45216;
 XX
 DT 02-NOV-1994 (first entry)
 XX
 DE Recombinant M24-M5-M6-M19 (linker variant) gene.
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKX223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX
 OS Synthetic.
 XX
 PN W09406421-A.
 XX
 PD 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 DR WPI; 1994-118122/14.
 DR P-PSDB; R50994.
 XX
 PT New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX
 PS Disclosure; Fig 6; 67pp; English.
 XX
 CC The sequences given in Q45214-23 encode hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC PKX223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and

PS Disclosure; Fig 6A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A

CC streptococcal M proteins comprising N-terminal peptide fragments of M

CC proteins that elicit opsonic antibodies against multiple serotypes

CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies

CC generated using the hybrid proteins are against one or more M protein

CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a

CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

CC method for immunising a mammal against streptococcal infections,

CC comprising administering an immunogenic multivalent hybrid M protein to

CC the mammal. The multivalent hybrid M proteins are useful for eliciting

CC opsonic or protective antibodies to the M proteins of different serotypes

CC of group A streptococci, and may therefore be used as vaccines to protect

CC against and control infection by type A streptococci. Type A streptococci

CC are not only responsible for streptococcal pharyngitis (strep throat),

CC forms of pneumonia and a condition resembling toxic shock, but are also

CC involved in the development of acute rheumatic fever (ARF) and rheumatic

CC heart disease. In a patient with ARF, antibodies formed during a group A

CC streptococcal infection are also cross-reactive with heart tissue, which

CC indicates that the streptococci and host tissue contain similar antigenic

CC motifs. The new multivalent vaccines are capable of raising sero-specific

CC antibodies against various serotypes of group A streptococci which are

CC not cross-reactive with human heart tissue. Sequences A57893-A57902

CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M

CC proteins generated in the disclosure of the invention.

XX

XX Sequence 822 BP; 332 A; 136 C; 182 G; 172 T; 0 other;

SQ

Query Match 18.7%; Score 217; DB 21; Length 822;

Best Local Similarity 100.0%; Pred. No. 3.4e-98;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 910 gtgcgactaggtctcagacagatactctggaaaaagatcacagaagctgtcgaagttt 969

Db 4 gtgcgactaggtctcagacagatactctggaaaaagatcacagaagctgtcgaagttt 63

Qy 970 gagatgaaaaacaatactctaaacttaagaatagtagtacttaagttttaataaagc 1029

Db 64 gagatgaaaaacaatactctaaacttaagaatagtagtacttaagttttaataaagc 123

Qy 1030 ttaaaagatcataatgatgagtttaactgaagagtgtagtaagtctaaagaaactacgt 1089

Db 124 ttaaaagatcataatgatgagtttaactgaagagtgtagtaagtctaaagaaactacgt 183

Qy 1090 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 1126

Db 184 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 220

RESULT 12

Q45221

ID Q45221 standard; DNA; 1029 BP.

XX

XX Q45221;

AC

XX

DT 02-NOV-1994 (first entry)

XX

DE Tetraivalent-C repeat gene.

XX

KW primer; PCR; amplify; polymerase chain reaction; construct; hybrid;

KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;

KW tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;

KW restriction enzyme site; multivalent M protein; immunisation; group A;

KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

KW antibody; heart tissue; antigen; serotype; mucosal; ss.

OS Synthetic.

XX

XX W09406421-A.

PN

XX

PD 31-MAR-1994.

XX

PF 15-SEP-1993; 93WO-US08703.

XX

PR 16-SEP-1992; 92US-0945954.

XX

PA (UYTE-) UNIV TENNESSEE RES CORP.

XX

PI Dale JB, Lederer JW;

XX

DR WPI: 1994-118122/14.

XX

DR P-PSDB: R50999.

XX

PT New immunogenic hybrid proteins derives from streptococcal M proteins

PT - induces opsonic antibodies, for protective immunisation against

PT against multiple group A streptococci serotypes

PS

XX Disclosure; Fig 11; 67pp; English.

XX The sequences given in Q45214-23 encode hybrid M proteins which

CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent

CC proteins were constructed using fragments of the 5' regions of emm

CC genes that were amplified by PCR, ligated in tandem and expressed in

CC pK223.3. The amplified regions pref. encode protective and not

CC tissue-cross-reactive epitopes, which can then be linked into one

CC protein molecule. The recombinant hybrid protein may contain 13

CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and

CC 35 from M19. Each section is linked by 2 amino acids specified by

CC the respective restriction enzyme sites that were synthesised into

CC the primers used to specify the PCR product. Multivalent M proteins

CC such as this may be used for protective immunisation against group A

CC streptococci, which esp. cause rheumatic fever and rheumatic heart

CC disease. Humoral antibodies raised against these proteins do not

CC react with heart tissue antigens but are effective against many

CC different serotypes. The multivalent proteins may also include

CC sequences which induce mucosal antibodies and do not require coupling

CC to an immunogenic carrier.

SQ

Sequence 1029 BP; 424 A; 168 C; 236 G; 211 T; 0 other;

Query Match 18.7%; Score 217; DB 15; Length 1029;

Best Local Similarity 100.0%; Pred. No. 3.4e-98;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 910 gtgcgactaggtctcagacagatactctggaaaaagatcacagaagctgtcgaagttt 969

Db 406 gtgcgactaggtctcagacagatactctggaaaaagatcacagaagctgtcgaagttt 465

Qy 970 gagatgaaaaacaatactctaaacttaagaatagtagtacttaagttttaataaagc 1029

Db 466 gagatgaaaaacaatactctaaacttaagaatagtagtacttaagttttaataaagc 525

Qy 1030 ttaaaagatcataatgatgagtttaactgaagagtgtagtaagtctaaagaaactacgt 1089

Db 526 ttaaaagatcataatgatgagtttaactgaagagtgtagtaagtctaaagaaactacgt 595

Qy 1090 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 1126

Db 586 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 622

RESULT 13

A57900

ID A57900 standard; DNA; 1029 BP.

XX

XX A57900;

AC

XX

DT 10-OCT-2000 (first entry)

XX

DE S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:16.

XX

KW Multivalent hybrid M protein; group A streptococcus; serotype;

Tue May 15 07:26:45 2001

immunogenic; sero-specific antibody; streptococcal infection;
cross reactivity; vaccine; acute rheumatic fever; ARF;
rheumatic heart disease; streptococcal pharyngitis; strep throat;
pneumonia; ds.

Streptococcus pyogenes.

Synthetic.

US6063386-A.

16-MAY-2000.

15-SEP-1997; 97US-0937271.

16-SEP-1992; 92US-0945954.

(UYTE-) UNIV TENNESSEE RES CORP.

Lederer JW, Dale JB;

WPI; 2000-364475/31.

P-PSDB; B03121.

New immunogenic recombinant hybrid M protein comprising amino-terminal
peptide fragments of streptococcal M protein useful as vaccine against
rheumatic fever and infections leading to rheumatic fever

Disclosure; Fig 11A-C; 52pp; English.

The invention relates to multivalent immunogenic hybrid group A
streptococcal M proteins comprising N-terminal peptide fragments of M
proteins that elicit opsonic antibodies against multiple serotypes
of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
generated using the hybrid proteins are against one or more M protein
serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
The invention also encompasses a recombinant DNA molecule comprising a
nucleotide sequence that encodes a multivalent hybrid M protein; and a
method for immunising a mammal against streptococcal infections,
comprising administering an immunogenic multivalent hybrid M protein to
the mammal. The multivalent hybrid M proteins are useful for eliciting
opsonic or protective antibodies to the M proteins of different serotypes
of group A streptococci, and may therefore be used as vaccines to protect
against and control infection by type A streptococci. Type A streptococci
are not only responsible for streptococcal pharyngitis (strep throat),
forms of pneumonia and a condition resembling toxic shock, but are also
involved in the development of acute rheumatic fever (ARF) and rheumatic
heart disease. In a patient with ARF, antibodies formed during a group A
streptococcal infection are also cross-reactive with heart tissue, which
indicates that the streptococci and host tissue contain similar antigenic
motifs. The new multivalent vaccines are capable of raising sero-specific
antibodies against various serotypes of group A streptococci which are
not cross-reactive with human heart tissue. Sequences A57893-A57902
represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
proteins generated in the disclosure of the invention.

Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 18.7%; Score 217; DB 21; Length 1029;
Best Local Similarity 100.0%; Pred. No. 3.4e-98;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 910 gtcgcgactaggcttcagacagactcttgataaacttaagtagtacttaagttttaataataaagcgc 1029
Db 406 gtcgcgactaggcttcagacagactcttgataaacttaagtagtacttaagttttaataataaagcgc 465
Qy 970 gagatagaataacacgttttaaaacttaagtagtacttaagttttaataataaagcgc 1029
Db 466 gagatagaataacacgttttaaaacttaagtagtacttaagttttaataataaagcgc 525
Qy 1030 ttaaaagatcataatgatgatgttaactgaagaggttgagtaagtgttaagaaactacgt 1089
Db 526 ttaaaagatcataatgatgatgttaactgaagaggttgagtaagtgttaagaaactacgt 585

Qy 1090 aaaaatgataaatacactatctgtgataaagctagtaaaa 1126
Db 586 aaaaatgataaatacactatctgtgataaagctagtaaaa 622

RESULT 14

T49317
ID T49317 standard; DNA; 1617 BP.

XX
AC T49317;

XX
DT 31-MAR-1997 (first entry)

XX
DE Type-6 M-protein emm6 gene.

XX
KW Type-6 M-protein; emm6 gene; Streptococcus pyogenes; signal peptide;
anchor peptide; C-terminal sorting signal; surface display;
fusion protein; protease-deficient; Streptococcus gordonii;
KW SPFX vector; fusion protein cleavage; spacer; protein secretion;
KW TEV-N1A protease; diagnostic; recombinant vaccine; therapy; ss.

XX
OS Streptococcus pyogenes.

XX
FH Key Location/Qualifiers

FT RBS 87..93

FT sig_peptide /*tag= a

FT /*tag= b

FT CDS 100..1551

FT /*tag= c

FT product Type-6 M-protein

FT mat_peptide 226..1548

FT /*tag= d

FT terminator 1577..1613

FT /*tag= e

FT /rpt_type= INVERTED

XX
PN W09640943-A1.

XX
PD 19-DEC-1996.

XX
PE 06-JUN-1996; 96WO-US09965.

XX
PX 07-JUN-1995; 95US-0472244.

XX
PA (UYRQ) UNIV ROCKEFELLER.

XX
PI Darzins A, Hraby D, Whitehead S;

XX
PX WPI; 1997-052336/05.

XX
DR P-PSDB; W08927.

XX
PT Expressing heterologous proteins in gram-positive bacteria -
PT produces proteins anchored to host's cell surface, or secreted from
PT cell

XX
PS Disclosure; Fig 3; 41pp; English.

XX
CC This sequence encodes Streptococcus pyogenes type-6 M-protein,
CC which is a cell wall surface protein. The N-terminal signal
CC peptide, part of the mature N-terminal sequence, the C-terminal
CC sorting signal (anchor) sequence and optionally a spacer region
CC from the protein may be expressed as a fusion with a target protein
CC in a protease-deficient gram-positive host, e.g. the human oral
CC commensal Streptococcus gordonii, in a new cloning method involving
CC the SPFX vector series (e.g. plasmid pSPFX1a). The resulting fusion
CC protein is expressed on the cell surface, and may be cleaved with a
CC protease (at a site adjacent to the anchor) for simple purification.
CC A vector without the anchor sequence may be used for protein
CC secretion. The method may be used in production of recombinant
CC diagnostic, vaccine and therapeutic proteins.

XX

SQ Sequence 1617 BP; 707 A; 261 C; 324 G; 325 T; 0 other;

Query Match 9.1%; Score 105; DB 18; Length 1617;
Best Local Similarity 100.0%; Pred. No. 1.4e-42;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 376 agagtgttcttcaggagcgtagaataaccggagacacgagagacttcttaacaag 435
|||||
DB 226 agagtgttcttcaggagcgtagaataaccggagacacgagagacttcttaacaag 435
|||||
QY 436 tatgacgtagagaactctatgtttacaagcacaagcagagagacttcttaacaag 480
|||||
DB 286 tatgacgtagagaactctatgtttacaagcacaagcagagagacttcttaacaag 330
|||||

RESULT 15
N50341
ID N50341 standard; DNA; 301 BP.
XX AC N50341;
XX 01-DEC-1991 (first entry)
DE Sequence of a portion of the emm6 gene that encodes the amino
terminal end of the M6 protein.
KW Fibrillar surface molecule; rheumatic fever; glomerulonephritis;
throat swabs; diagnosis; ss.
XX Streptococcus pyogenes (Group A streptococcus) strain D471.
FH Key Location/Qualifiers
FT CDS 1..126
FT /*tag= a
FT CDS 127..300
FT /*tag= b
FT /*product= N-terminal of M6
XX W08500832-A.
XX 28-FEB-1985.
XX 09-AUG-1984; 84WO-4001261.
XX 18-JUN-1984; 84US-0621716.
XX 10-AUG-1983; 83US-0521962.
XX (UYRO-) ROCKEFELLER UNIV.
XX (UYEM-) EMORY UNIV.
XX (UYEM) EMORY UNIV.
XX Scott JR, Fischetti VA;
XX WPI; 1985-062291/10.
XX P-PSDB; P50295.
XX Immunogenic streptococcal polypeptide prodn. - by recombinant DNA
methods, useful as vaccines and as diagnostic probes
XX Example; Fig 4; 46pp; English.
XX The inventors claim a polypeptide having an immunoreactive and
antigenic determinant of a Streptococcus pyogenes M protein produced
by a unicellular organism, and a DNA sequence encoding it. Also
claimed are purified DNA probes able to bind to the S.pyogenes gene
which codes for M protein. The polypeptides can be used in vaccines.
XX Sequence 301 BP; 125 A; 45 C; 55 G; 70 T; 6 other;

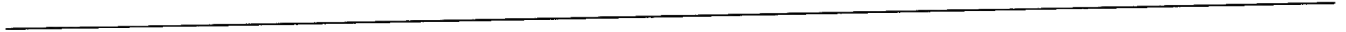
Query Match 5.4%; Score 63; DB 6; Length 301;
Best Local Similarity 100.0%; Pred. No. 1.1e-21;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 376 agagtgttcttcaggagcgtagaataaccggagacacgagagacttcttaacaag 435
|||||
DB 127 agagtgttcttcaggagcgtagaataaccggagacacgagagacttcttaacaag 186
|||||
QY 436 tat 438
|||||
DB 187 tat 189

Search completed: May 13, 2001, 06:15:08
Job time: 4558 sec

us-09-151-409-15.oligo.rng

Tue May 15 07:26:45 2001



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 05:04:00 ; Search time 66.69 seconds
(without alignments)
3031.788 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacgatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 302621 seqs, 87301344 residues

Word size : 10

Total number of hits satisfying chosen parameters: 58178

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	402	34.7	741	3	US-08-937-271-3
2	240	20.7	861	3	US-08-937-271-1
3	217	18.7	522	3	US-08-937-271-14
4	217	18.7	741	3	US-08-937-271-12
5	217	18.7	822	3	US-08-937-271-5
6	217	18.7	1029	3	US-08-937-271-16
7	151	13.0	1332	2	US-08-795-475-5
8	53	4.6	561	3	US-08-937-271-7
9	52	4.5	204	3	US-08-937-271-19
10	52	4.5	408	3	US-08-937-271-21
11	52	4.5	918	3	US-08-937-271-9
12	38	3.3	39	3	US-08-937-271-26
13	34	2.9	47	3	US-08-937-271-28
14	32	2.8	54	3	US-08-937-271-38
15	25	2.2	30	3	US-08-937-271-34
16	24	2.1	30	3	US-08-937-271-32
17	24	2.1	30	3	US-08-937-271-36
18	24	2.1	30	3	US-08-937-271-37
19	22	1.9	29	3	US-08-937-271-24
20	22	1.9	29	3	US-08-937-271-30
21	19	1.6	30	3	US-08-937-271-39
22	17	1.5	5077	2	US-08-687-956A-22
23	16	1.4	342	1	US-08-486-013-56
24	16	1.4	342	2	US-08-482-279-56
25	16	1.4	342	2	US-08-342-268-56
26	16	1.4	342	3	US-08-015-968-56
27	16	1.4	352	1	US-08-486-013-50
					Sequence 3, Appli
					Sequence 1, Appli
					Sequence 14, Appli
					Sequence 12, Appli
					Sequence 5, Appli
					Sequence 16, Appli
					Sequence 5, Appli
					Sequence 7, Appli
					Sequence 19, Appli
					Sequence 9, Appli
					Sequence 26, Appli
					Sequence 28, Appli
					Sequence 38, Appli
					Sequence 34, Appli
					Sequence 32, Appli
					Sequence 37, Appli
					Sequence 21, Appli
					Sequence 20, Appli
					Sequence 39, Appli
					Sequence 39, Appli
					Sequence 22, Appli
					Sequence 56, Appli
					Sequence 56, Appli
					Sequence 56, Appli
					Sequence 56, Appli
					Sequence 50, Appli

1.4	352	2	US-08-482-279-50	Sequence 50, Appli
1.4	352	2	US-08-342-268-50	Sequence 50, Appli
1.4	352	3	US-09-015-968-50	Sequence 50, Appli
1.4	375	1	US-08-486-013-49	Sequence 49, Appli
1.4	375	2	US-08-482-279-49	Sequence 49, Appli
1.4	375	3	US-08-342-268-49	Sequence 49, Appli
1.4	375	3	US-09-015-968-49	Sequence 49, Appli
1.4	377	1	US-08-486-013-48	Sequence 48, Appli
1.4	377	2	US-08-482-279-48	Sequence 48, Appli
1.4	377	3	US-08-342-268-48	Sequence 48, Appli
1.4	401	1	US-09-015-968-48	Sequence 48, Appli
1.4	401	2	US-08-482-279-54	Sequence 54, Appli
1.4	401	3	US-08-342-268-54	Sequence 54, Appli
1.4	419	1	US-09-015-968-54	Sequence 54, Appli
1.4	419	1	US-08-486-013-58	Sequence 58, Appli
1.4	419	1	US-08-486-013-62	Sequence 62, Appli
1.4	419	2	US-08-486-013-63	Sequence 63, Appli
1.4	419	2	US-08-482-279-58	Sequence 58, Appli
1.4	419	2	US-08-482-279-62	Sequence 62, Appli
1.4	419	2	US-08-482-279-63	Sequence 63, Appli
1.4	419	2	US-08-342-268-58	Sequence 58, Appli
1.4	419	2	US-08-342-268-62	Sequence 62, Appli
1.4	419	3	US-09-015-968-58	Sequence 62, Appli
1.4	419	3	US-09-015-968-62	Sequence 62, Appli
1.4	422	1	US-08-486-013-40	Sequence 40, Appli
1.4	422	1	US-08-486-013-45	Sequence 45, Appli
1.4	422	2	US-08-482-279-45	Sequence 45, Appli
1.4	422	2	US-08-342-268-45	Sequence 45, Appli
1.4	422	3	US-09-015-968-45	Sequence 45, Appli
1.4	473	1	US-08-480-784-23	Sequence 23, Appli
1.4	473	1	US-08-483-553-23	Sequence 23, Appli
1.4	473	1	US-08-487-002-23	Sequence 23, Appli
1.4	473	1	US-08-483-554B-23	Sequence 23, Appli
1.4	473	1	US-08-488-011B-23	Sequence 23, Appli
1.4	473	4	US-08-850-727-23	Sequence 23, Appli
1.4	473	5	PCT-US95-10202-23	Sequence 23, Appli
1.4	473	5	PCT-US95-10203-23	Sequence 23, Appli
1.4	577	1	US-08-318-905-16	Sequence 16, Appli
1.4	577	1	US-08-483-232-16	Sequence 16, Appli
1.4	577	2	US-08-483-140-16	Sequence 16, Appli
1.4	577	2	US-08-485-938A-16	Sequence 16, Appli
1.4	577	3	US-09-328-474-16	Sequence 16, Appli
1.4	577	3	US-09-100-546-16	Sequence 16, Appli
1.4	577	4	US-09-010-715-16	Sequence 16, Appli
1.4	577	4	US-09-577-758-16	Sequence 16, Appli
1.4	1335	1	US-08-483-232-30	Sequence 30, Appli
1.4	1335	2	US-08-485-938A-30	Sequence 30, Appli
1.4	1335	3	US-09-328-474-30	Sequence 30, Appli
1.4	1335	3	US-09-100-546-30	Sequence 30, Appli
1.4	1335	4	US-09-010-715-30	Sequence 30, Appli
1.4	1335	4	US-09-577-758-30	Sequence 30, Appli
1.4	1361	2	US-08-387-858A-9	Sequence 9, Appli
1.4	1361	2	US-09-294-384B-9	Sequence 9, Appli
1.4	1520	1	US-08-470-187-7	Sequence 7, Appli
1.4	1520	1	US-08-318-905-7	Sequence 7, Appli
1.4	1520	1	US-08-483-232-7	Sequence 7, Appli
1.4	1520	2	US-08-483-140-7	Sequence 7, Appli
1.4	1520	2	US-08-485-938A-7	Sequence 7, Appli
1.4	1520	3	US-09-328-474-7	Sequence 7, Appli
1.4	1520	3	US-09-100-546-7	Sequence 7, Appli
1.4	1520	4	US-09-010-715-7	Sequence 7, Appli
1.4	1520	4	US-09-577-758-7	Sequence 7, Appli

c 539	14	1.2	2057	3	US-09-055-699-42	Sequence 42, Appl	c 612	14	1.2	2659	3	US-08-749-522-1	Sequence 1, Appl
c 540	14	1.2	2057	4	US-09-273-565-42	Sequence 42, Appl	c 613	14	1.2	2662	1	US-08-451-715A-5	Sequence 5, Appl
c 541	14	1.2	2074	1	US-08-414-685-1	Sequence 1, Appl	614	14	1.2	2679	1	US-07-977-434-11	Sequence 11, Appl
c 542	14	1.2	2075	3	US-08-262-220-3	Sequence 3, Appl	615	14	1.2	2679	1	US-08-458-819-11	Sequence 11, Appl
c 543	14	1.2	2075	3	US-08-471-723-3	Sequence 3, Appl	616	14	1.2	2679	5	PCT-US91-07035-11	Sequence 5, Appl
c 544	14	1.2	2075	3	US-08-468-878-3	Sequence 3, Appl	c 617	14	1.2	2700	1	US-08-484-105-5	Sequence 5, Appl
c 545	14	1.2	2075	4	US-08-750-494-3	Sequence 7, Appl	c 618	14	1.2	2700	1	US-08-484-106-5	Sequence 5, Appl
c 546	14	1.2	2081	2	US-09-096-982-7	Sequence 7, Appl	c 619	14	1.2	2755	3	US-08-749-522-2	Sequence 2, Appl
c 547	14	1.2	2081	2	US-08-653-650A-7	Sequence 7, Appl	c 620	14	1.2	2809	1	US-08-484-105-3	Sequence 3, Appl
c 548	14	1.2	2086	3	US-08-724-586-1	Sequence 1, Appl	621	14	1.2	2809	1	US-08-484-106-3	Sequence 3, Appl
c 549	14	1.2	2089	1	US-08-552-142A-1	Sequence 1, Appl	622	14	1.2	2848	3	US-08-539-205A-3	Sequence 3, Appl
c 550	14	1.2	2089	5	PCT-US95-05741-1	Sequence 1, Appl	623	14	1.2	2859	2	US-08-506-340A-2	Sequence 2, Appl
c 551	14	1.2	2089	5	PCT-US95-05741-1	Sequence 1, Appl	624	14	1.2	2946	3	US-08-968-563-6	Sequence 3, Appl
c 552	14	1.2	2117	2	US-08-655-640-1	Sequence 1, Appl	625	14	1.2	2946	4	US-08-969-683A-6	Sequence 6, Appl
c 553	14	1.2	2124	2	US-08-655-640-3	Sequence 3, Appl	626	14	1.2	2970	3	US-08-974-180-14	Sequence 14, Appl
c 554	14	1.2	2157	2	US-08-132-990A-7	Sequence 7, Appl	c 627	14	1.2	2970	3	US-08-974-180-14	Sequence 14, Appl
c 555	14	1.2	2157	5	PCT-US92-09382-7	Sequence 7, Appl	628	14	1.2	2974	1	US-08-290-978A-4	Sequence 4, Appl
c 556	14	1.2	2178	3	US-08-781-891-72	Sequence 72, Appl	629	14	1.2	2974	2	US-08-780-869-4	Sequence 4, Appl
c 557	14	1.2	2236	2	US-08-484-933B-13	Sequence 13, Appl	630	14	1.2	3011	1	US-07-821-716-1	Sequence 1, Appl
c 558	14	1.2	2236	2	US-08-484-158B-13	Sequence 13, Appl	631	14	1.2	3030	2	US-08-680-326-24	Sequence 24, Appl
c 559	14	1.2	2236	2	US-08-484-596A-13	Sequence 13, Appl	632	14	1.2	3068	1	US-07-984-044A-3	Sequence 3, Appl
c 560	14	1.2	2236	2	US-08-480-150A-13	Sequence 13, Appl	c 633	14	1.2	3068	1	US-08-458-393-3	Sequence 3, Appl
c 561	14	1.2	2236	3	US-08-458-731-13	Sequence 13, Appl	634	14	1.2	3074	4	US-09-087-277-1	Sequence 1, Appl
c 562	14	1.2	2236	3	US-08-149-223A-13	Sequence 13, Appl	635	14	1.2	3111	3	US-08-891-845-1	Sequence 1, Appl
c 563	14	1.2	2258	1	US-07-720-589-1	Sequence 1, Appl	c 636	14	1.2	3116	1	US-08-149-103-2	Sequence 2, Appl
c 564	14	1.2	2258	1	US-07-720-589-1	Sequence 1, Appl	c 637	14	1.2	3116	1	US-08-451-883-2	Sequence 2, Appl
c 565	14	1.2	2258	5	PCT-US92-05539-1	Sequence 5, Appl	c 638	14	1.2	3291	5	PCT-US96-03940-10	Sequence 10, Appl
c 566	14	1.2	2263	2	US-08-687-865A-1	Sequence 1, Appl	c 639	14	1.2	3330	1	US-08-149-103-1	Sequence 1, Appl
c 567	14	1.2	2288	2	US-08-568-459A-5	Sequence 5, Appl	c 640	14	1.2	3330	1	US-08-451-883-1	Sequence 1, Appl
c 568	14	1.2	2288	2	US-08-487-826B-5	Sequence 5, Appl	641	14	1.2	3337	1	US-08-072-610-1	Sequence 1, Appl
c 569	14	1.2	2290	4	US-08-961-083-65	Sequence 65, Appl	642	14	1.2	3337	2	US-08-719-822B-1	Sequence 1, Appl
c 570	14	1.2	2295	1	US-08-375-300-3	Sequence 3, Appl	643	14	1.2	3426	1	US-08-234-939-1	Sequence 1, Appl
c 571	14	1.2	2295	1	US-09-177-431-3	Sequence 3, Appl	644	14	1.2	3426	1	US-08-558-865-1	Sequence 1, Appl
c 572	14	1.2	2295	5	PCT-US95-16930-3	Sequence 3, Appl	645	14	1.2	3426	3	US-08-654-025-6	Sequence 6, Appl
c 573	14	1.2	2308	4	US-09-377-155-1	Sequence 4, Appl	646	14	1.2	3466	1	US-08-468-036-38	Sequence 38, Appl
c 574	14	1.2	2308	1	US-08-154-019-1	Sequence 1, Appl	647	14	1.2	3466	2	US-08-293-728-1	Sequence 38, Appl
c 575	14	1.2	2319	1	US-08-461-333-1	Sequence 1, Appl	c 648	14	1.2	3498	3	US-08-376-843-38	Sequence 38, Appl
c 576	14	1.2	2319	3	US-08-464-167-1	Sequence 1, Appl	c 649	14	1.2	3498	4	US-09-421-868-1	Sequence 1, Appl
c 577	14	1.2	2319	3	US-09-158-313-1	Sequence 1, Appl	c 650	14	1.2	3632	1	US-08-424-788-4	Sequence 1, Appl
c 578	14	1.2	2319	4	US-08-476-798-1	Sequence 1, Appl	651	14	1.2	3632	2	US-08-110-683-1	Sequence 1, Appl
c 579	14	1.2	2325	3	US-08-902-632-3	Sequence 3, Appl	652	14	1.2	3632	2	US-08-683-743-1	Sequence 1, Appl
c 580	14	1.2	2360	1	US-08-145-681-1	Sequence 1, Appl	653	14	1.2	3632	2	US-08-477-166-1	Sequence 1, Appl
c 581	14	1.2	2360	1	US-08-250-308-1	Sequence 1, Appl	654	14	1.2	3632	2	US-08-472-097-1	Sequence 1, Appl
c 582	14	1.2	2360	1	US-08-453-703-1	Sequence 1, Appl	655	14	1.2	3632	5	PCT-US93-11638-1	Sequence 1, Appl
c 583	14	1.2	2360	1	US-08-039-364-1	Sequence 1, Appl	656	14	1.2	3642	3	US-08-946-026-16	Sequence 16, Appl
c 584	14	1.2	2360	2	US-08-456-106-1	Sequence 1, Appl	c 657	14	1.2	3656	4	US-08-393-734-1	Sequence 1, Appl
c 585	14	1.2	2360	4	US-08-456-108-1	Sequence 1, Appl	c 658	14	1.2	3656	4	US-08-894-489-1	Sequence 1, Appl
c 586	14	1.2	2360	5	PCT-US93-03614-1	Sequence 1, Appl	c 659	14	1.2	3684	1	US-08-448-170-7	Sequence 7, Appl
c 587	14	1.2	2363	1	US-08-096-947-2	Sequence 2, Appl	c 660	14	1.2	3684	1	US-08-961-803-5	Sequence 5, Appl
c 588	14	1.2	2363	1	US-07-919-140B-2	Sequence 2, Appl	c 661	14	1.2	3713	1	US-08-330-537-3	Sequence 3, Appl
c 589	14	1.2	2363	5	PCT-US93-06939-2	Sequence 2, Appl	c 662	14	1.2	3713	1	US-08-330-537-4	Sequence 4, Appl
c 590	14	1.2	2368	3	US-08-714-918-2	Sequence 2, Appl	c 663	14	1.2	3765	4	US-07-705-490-1	Sequence 1, Appl
c 591	14	1.2	2368	4	US-09-265-315-2	Sequence 2, Appl	c 664	14	1.2	3765	4	US-07-751-891B-1	Sequence 1, Appl
c 592	14	1.2	2368	4	US-09-265-315-2	Sequence 2, Appl	665	14	1.2	3812	1	US-08-188-582-19	Sequence 19, Appl
c 593	14	1.2	2369	4	US-08-910-923-2	Sequence 2, Appl	666	14	1.2	3812	1	US-08-646-715-19	Sequence 19, Appl
c 594	14	1.2	2404	1	US-08-484-105-7	Sequence 7, Appl	667	14	1.2	3848	2	US-08-808-931-14	Sequence 14, Appl
c 595	14	1.2	2404	1	US-08-484-106-7	Sequence 7, Appl	668	14	1.2	3848	3	US-08-808-323-14	Sequence 14, Appl
c 596	14	1.2	2411	1	US-07-952-755-1	Sequence 1, Appl	669	14	1.2	3848	3	US-09-050-603A-14	Sequence 14, Appl
c 597	14	1.2	2411	1	US-08-443-679-1	Sequence 1, Appl	670	14	1.2	3848	3	US-09-102-420B-14	Sequence 14, Appl
c 598	14	1.2	2426	5	PCT-US91-09422-20	Sequence 20, Appl	c 671	14	1.2	3891	2	US-09-112-096-28	Sequence 28, Appl
c 599	14	1.2	2454	1	US-08-359-696-3	Sequence 3, Appl	c 672	14	1.2	3998	4	US-08-966-630-3	Sequence 3, Appl
c 600	14	1.2	2463	1	US-08-307-444A-10	Sequence 10, Appl	c 673	14	1.2	3998	4	US-09-066-046-5	Sequence 5, Appl
c 601	14	1.2	2463	1	US-08-587-389-10	Sequence 10, Appl	c 674	14	1.2	4010	2	US-08-785-310A-3	Sequence 3, Appl
c 602	14	1.2	2467	2	US-08-701-240-1	Sequence 1, Appl	c 675	14	1.2	4052	1	US-08-833-226-1	Sequence 1, Appl
c 603	14	1.2	2467	4	US-09-138-236-1	Sequence 1, Appl	c 676	14	1.2	4080	1	US-08-375-300-1	Sequence 1, Appl
c 604	14	1.2	2522	3	US-08-714-918-93	Sequence 93, Appl	c 677	14	1.2	4080	3	US-09-177-431-1	Sequence 1, Appl
c 605	14	1.2	2522	4	US-09-265-315-93	Sequence 93, Appl	c 678	14	1.2	4080	5	PCT-US95-16930-1	Sequence 1, Appl
c 606	14	1.2	2522	4	US-09-265-315-93	Sequence 93, Appl	c 679	14	1.2	4095	5	PCT-US91-09422-18	Sequence 18, Appl
c 607	14	1.2	2547	3	US-08-262-220-7	Sequence 7, Appl	680	14	1.2	4112	1	US-08-340-203A-2	Sequence 2, Appl
c 608	14	1.2	2547	3	US-08-471-733-7	Sequence 7, Appl	681	14	1.2	4112	2	US-08-452-567-2	Sequence 2, Appl
c 609	14	1.2	2547	3	US-08-468-878-7	Sequence 7, Appl	682	14	1.2	4112	2	US-08-452-427-2	Sequence 2, Appl
c 610	14	1.2	2547	4	US-08-750-494-7	Sequence 7, Appl	683	14	1.2	4112	4	US-09-085-407-2	Sequence 2, Appl
c 611	14	1.2	2559	2	US-08-248-839C-10	Sequence 10, Appl	c 684	14	1.2	4140	3	US-08-894-731-2	Sequence 2, Appl

c 685	14	1.2	4195	1	US-08-200-016-1	Sequence 1, Appl	14	1.2	7351	2	US-08-458-356-127	Sequence 127, App
c 686	14	1.2	4223	4	US-08-845-258-7	Sequence 7, Appl	14	1.2	7351	2	US-08-658-665-39	Sequence 39, Appl
c 687	14	1.2	4223	4	US-08-845-258-7	Sequence 45, Appl	14	1.2	7351	4	US-08-796-101-3	Sequence 3, Appl
c 688	14	1.2	4286	2	US-08-249-617-1	Sequence 1, Appl	14	1.2	7705	2	US-08-687-080-115	Sequence 115, App
c 689	14	1.2	4286	4	US-08-634-350-1	Sequence 1, Appl	14	1.2	8313	1	US-08-232-463-2	Sequence 2, Appl
c 690	14	1.2	4286	4	US-09-413-304-7	Sequence 7, Appl	14	1.2	8313	3	US-08-680-506-1	Sequence 1, Appl
c 691	14	1.2	4322	1	US-08-673-789-1	Sequence 1, Appl	14	1.2	8321	3	US-08-716-351A-1	Sequence 1, Appl
c 692	14	1.2	4362	2	US-08-455-073A-1	Sequence 1, Appl	14	1.2	8535	3	US-08-815-809-7	Sequence 7, Appl
c 693	14	1.2	4434	3	US-08-815-809-1	Sequence 1, Appl	14	1.2	8535	3	US-08-232-463-5	Sequence 5, Appl
c 694	14	1.2	4435	2	US-08-792-824-1	Sequence 3, Appl	14	1.2	8775	1	US-08-815-809-7	Sequence 7, Appl
c 695	14	1.2	4472	2	US-08-816-155B-3	Sequence 3, Appl	14	1.2	8868	3	US-08-232-463-3	Sequence 3, Appl
c 696	14	1.2	4472	4	US-07-718-575-11	Sequence 11, Appl	14	1.2	9454	1	US-08-815-809-7	Sequence 7, Appl
c 697	14	1.2	4608	1	US-08-481-206-11	Sequence 11, Appl	14	1.2	9454	1	US-08-232-463-4	Sequence 4, Appl
c 698	14	1.2	4608	2	US-08-486-269A-11	Sequence 1, Appl	14	1.2	9997	1	US-08-246-982A-15	Sequence 15, Appl
c 699	14	1.2	4616	1	US-08-340-203A-1	Sequence 1, Appl	14	1.2	9997	1	US-08-453-265-15	Sequence 15, Appl
c 700	14	1.2	4616	2	US-08-452-567-1	Sequence 1, Appl	14	1.2	10103	2	US-08-457-273B-7	Sequence 7, Appl
c 701	14	1.2	4616	2	US-08-452-427-1	Sequence 1, Appl	14	1.2	10103	2	US-08-920-812-2	Sequence 2, Appl
c 702	14	1.2	4616	2	US-09-085-407-1	Sequence 5, Appl	14	1.2	10207	1	US-08-920-812-2	Sequence 2, Appl
c 703	14	1.2	4616	4	US-08-867-0308-5	Sequence 5, Appl	14	1.2	10207	1	US-08-921-177-2	Sequence 2, Appl
c 704	14	1.2	4951	2	PCT-US95-06119-5	Sequence 5, Appl	14	1.2	10207	1	US-08-362-577C-2	Sequence 2, Appl
c 705	14	1.2	4951	5	US-08-920-827-18	Sequence 18, Appl	14	1.2	10207	2	US-08-920-828-2	Sequence 2, Appl
c 706	14	1.2	4954	1	US-08-920-827-18	Sequence 18, Appl	14	1.2	10747	2	US-08-147-777-1	Sequence 1, Appl
c 707	14	1.2	4954	1	US-08-921-177-18	Sequence 18, Appl	14	1.2	10747	2	US-08-452-872-1	Sequence 1, Appl
c 708	14	1.2	4954	1	US-08-362-577C-18	Sequence 18, Appl	14	1.2	10747	5	PCT-US93-03985-1	Sequence 5, Appl
c 709	14	1.2	4954	1	US-08-920-828-18	Sequence 18, Appl	14	1.2	10970	3	US-08-716-351A-5	Sequence 5, Appl
c 710	14	1.2	4954	2	US-08-232-463-1	Sequence 18, Appl	14	1.2	11225	6	5182210-9	Patent No. 5182210
c 711	14	1.2	4997	1	US-08-680-506-2	Sequence 18, Appl	14	1.2	11225	6	US-08-981-803-29	Sequence 29, Appl
c 712	14	1.2	5027	3	US-08-870-518-7	Sequence 18, Appl	14	1.2	11478	4	US-08-816-155B-8	Sequence 8, Appl
c 713	14	1.2	5183	2	US-08-870-518-7	Sequence 18, Appl	14	1.2	11478	4	US-09-079-587-8	Sequence 8, Appl
c 714	14	1.2	5183	2	US-08-870-518-7	Sequence 18, Appl	14	1.2	19877	4	US-09-053-197A-3	Sequence 3, Appl
c 715	14	1.2	5208	3	US-08-781-891-70	Sequence 70, Appl	14	1.2	22108	3	US-08-147-777-3	Sequence 3, Appl
c 716	14	1.2	5208	6	5453363-1	Patent No. 5453363	14	1.2	22108	3	US-08-452-872-3	Sequence 3, Appl
c 717	14	1.2	5253	2	US-08-290-731C-3	Sequence 3, Appl	14	1.2	24979	5	PCT-US93-03985-3	Sequence 3, Appl
c 718	14	1.2	5253	3	US-08-073-354-6	Sequence 6, Appl	14	1.2	24979	5	US-08-884-324-14	Sequence 14, Appl
c 719	14	1.2	5339	3	US-08-656-005A-6	Sequence 6, Appl	14	1.2	28994	3	US-08-884-324-14	Sequence 14, Appl
c 720	14	1.2	5339	4	US-09-073-259-6	Sequence 6, Appl	14	1.2	40138	4	US-09-090-793-12	Sequence 12, Appl
c 721	14	1.2	5339	4	US-09-363-095-6	Sequence 6, Appl	14	1.2	43676	4	US-09-356-952-12	Sequence 12, Appl
c 722	14	1.2	5342	3	US-08-902-632-1	Sequence 1, Appl	14	1.2	51952	4	US-08-947-823-1	Sequence 1, Appl
c 723	14	1.2	5342	3	US-08-902-632-1	Sequence 5, Appl	14	1.2	51952	4	US-09-128-155-16	Sequence 16, Appl
c 724	14	1.2	5342	3	US-08-656-005A-5	Sequence 5, Appl	14	1.2	152331	4	US-09-128-155-16	Sequence 16, Appl
c 725	14	1.2	5342	4	US-09-073-259-5	Sequence 5, Appl	14	1.2	152331	4	US-09-128-155-17	Sequence 17, Appl
c 726	14	1.2	5342	4	US-09-363-095-5	Sequence 5, Appl	14	1.2	152331	4	US-08-724-394A-20	Sequence 20, Appl
c 727	14	1.2	5433	4	US-08-929-329-1	Sequence 1, Appl	14	1.2	248240	2	US-08-724-394A-21	Sequence 21, Appl
c 728	14	1.2	5433	5	PCT-US96-03940-1	Sequence 1, Appl	14	1.2	248240	2	US-08-292-620A-453	Sequence 453, App
c 729	14	1.2	5668	4	US-09-112-096-14	Sequence 14, Appl	13	1.1	15	2	US-08-292-620A-553	Sequence 553, App
c 730	14	1.2	5977	3	US-09-024-020B-1	Sequence 1, Appl	13	1.1	15	2	US-08-292-620A-553	Sequence 553, App
c 731	14	1.2	6007	3	US-09-024-020B-2	Sequence 2, Appl	13	1.1	15	4	US-09-071-845-453	Sequence 453, App
c 732	14	1.2	6030	1	US-08-441-139-8	Sequence 8, Appl	13	1.1	15	4	US-09-071-845-460	Sequence 460, App
c 733	14	1.2	6312	4	US-09-011-745-7	Sequence 7, Appl	13	1.1	15	4	US-09-071-845-553	Sequence 553, App
c 734	14	1.2	6356	3	US-09-024-020B-7	Sequence 43, Appl	13	1.1	17	1	US-08-373-124A-922	Sequence 922, App
c 735	14	1.2	6586	1	US-08-727-034-1	Sequence 1, Appl	13	1.1	17	1	US-08-373-124A-924	Sequence 924, App
c 736	14	1.2	6639	2	US-08-147-777-2	Sequence 2, Appl	13	1.1	17	1	US-08-435-628-922	Sequence 922, App
c 737	14	1.2	6693	3	US-08-452-872-2	Sequence 2, Appl	13	1.1	17	1	US-08-435-628-924	Sequence 924, App
c 738	14	1.2	6693	3	PCT-US93-03985-2	Sequence 8, Appl	13	1.1	17	1	US-08-435-628-924	Sequence 924, App
c 739	14	1.2	6826	5	US-09-024-020B-8	Sequence 2, Appl	13	1.1	17	2	US-08-435-628-926	Sequence 926, App
c 740	14	1.2	6961	1	US-08-727-034-2	Sequence 1, Appl	13	1.1	17	2	US-08-292-620A-1735	Sequence 1735, Ap
c 741	14	1.2	7001	1	US-08-258-261B-1	Sequence 1, Appl	13	1.1	18	2	US-09-071-845-1735	Sequence 26, Appl
c 742	14	1.2	7001	1	US-08-456-837-1	Sequence 1, Appl	13	1.1	18	2	US-09-166-203-26	Sequence 80, Appl
c 743	14	1.2	7001	1	US-08-457-342-1	Sequence 1, Appl	13	1.1	18	2	US-09-166-203-26	Sequence 118, App
c 744	14	1.2	7001	1	US-08-457-646A-1	Sequence 1, Appl	13	1.1	20	3	US-09-166-186-118	Sequence 312, App
c 745	14	1.2	7001	1	US-08-458-076A-1	Sequence 1, Appl	13	1.1	20	3	US-09-009-913-312	Sequence 339, App
c 746	14	1.2	7001	1	US-08-457-335A-1	Sequence 1, Appl	13	1.1	22	2	US-08-117-952-339	Sequence 7, Appl
c 747	14	1.2	7001	1	US-08-761-258-6	Sequence 6, Appl	13	1.1	22	4	US-08-853-948B-7	Sequence 6, Appl
c 748	14	1.2	7001	1	US-08-729-214-1	Sequence 1, Appl	13	1.1	23	1	US-08-401-512-6	Sequence 5, Appl
c 749	14	1.2	7001	2	US-09-977-306-6	Sequence 1, Appl	13	1.1	23	1	US-08-376-296-5	Sequence 5, Appl
c 750	14	1.2	7001	4	US-08-978-762-46	Sequence 46, Appl	13	1.1	23	3	US-08-910-691-5	Sequence 32, Appl
c 751	14	1.2	7091	4	US-08-375-313-26	Sequence 26, Appl	13	1.1	26	2	US-08-520-678A-32	Sequence 32, Appl
c 752	14	1.2	7244	1	US-08-224-391-83	Sequence 83, Appl	13	1.1	27	3	US-08-986-331-1	Sequence 54, Appl
c 753	14	1.2	7351	1	US-08-484-304-83	Sequence 83, Appl	13	1.1	28	1	US-08-485-603-54	Sequence 54, Appl
c 754	14	1.2	7351	1	US-08-484-304-83	Sequence 83, Appl	13	1.1	28	1	US-08-757-180-53	Sequence 57, Appl
c 755	14	1.2	7351	2	US-08-184-009-127	Sequence 127, App	13	1.1	28	1	US-08-745-638-54	Sequence 54, Appl
c 756	14	1.2	7351	2	US-08-566-398-39	Sequence 39, Appl	13	1.1	29	1	US-08-295-676A-16	Sequence 16, Appl
c 757	14	1.2	7351	2	US-08-566-398-39	Sequence 39, Appl	13	1.1	29	1	US-08-295-676A-16	Sequence 16, Appl

c 831	13	1.1	29	2	US-08-948-591-16	Sequence 16, Appl	c 904	13	1.1	183	3	US-08-450-653-2	Sequence 2, Appl
c 832	13	1.1	30	1	US-08-605-541B-3	Sequence 3, Appl	c 905	13	1.1	183	4	US-09-014-416-60	Sequence 60, Appl
c 833	13	1.1	31	2	US-08-629-001A-25	Sequence 25, Appl	c 906	13	1.1	183	5	PCT-US94-09653A-30	Sequence 30, Appl
c 834	13	1.1	32	3	US-08-642-274D-104	Sequence 104, App	c 907	13	1.1	192	3	US-08-416-711-5	Sequence 5, Appl
c 835	13	1.1	31	4	US-08-960-780-66	Sequence 66, Appl	c 908	13	1.1	192	2	US-08-520-678A-31	Sequence 31, Appl
c 836	13	1.1	31	4	US-08-960-780-66	Sequence 114, App	c 909	13	1.1	200	4	US-09-014-416-64	Sequence 64, Appl
c 837	13	1.1	32	1	US-08-118-469A-10	Sequence 10, Appl	c 910	13	1.1	222	2	US-08-809-267-1	Sequence 1, Appl
c 838	13	1.1	32	1	US-08-909-119-10	Sequence 10, Appl	c 911	13	1.1	222	2	PCT-US95-1366A-1	Sequence 1, Appl
c 839	13	1.1	33	3	US-09-129-740-1	Sequence 1, Appl	c 912	13	1.1	224	2	US-08-731-272A-26	Sequence 26, Appl
c 840	13	1.1	34	2	US-08-788-943A-22	Sequence 22, Appl	c 913	13	1.1	227	2	US-08-520-678A-28	Sequence 28, Appl
c 841	13	1.1	34	2	US-08-534-685-2	Sequence 46, Appl	c 914	13	1.1	239	1	US-08-090-523-50	Sequence 50, Appl
c 842	13	1.1	35	1	US-08-476-008-46	Sequence 46, Appl	c 915	13	1.1	239	1	US-08-398-627-50	Sequence 50, Appl
c 843	13	1.1	35	1	US-08-306-063-46	Sequence 46, Appl	c 916	13	1.1	242	3	US-08-482-304-15	Sequence 15, Appl
c 844	13	1.1	35	1	US-08-833-485-46	Sequence 46, Appl	c 917	13	1.1	242	3	US-08-483-474-15	Sequence 15, Appl
c 845	13	1.1	36	1	US-08-196-538-13	Sequence 13, Appl	c 918	13	1.1	248	2	US-07-594-921C-10	Sequence 10, Appl
c 846	13	1.1	36	3	US-09-129-740-4	Sequence 4, Appl	c 919	13	1.1	250	3	US-09-118-841B-6	Sequence 6, Appl
c 847	13	1.1	38	1	US-07-829-954-3	Sequence 3, Appl	c 920	13	1.1	250	4	US-09-172-108-38	Sequence 38, Appl
c 848	13	1.1	39	4	US-09-014-416-10	Sequence 10, Appl	c 921	13	1.1	250	4	US-09-414-439-6	Sequence 6, Appl
c 849	13	1.1	40	1	US-08-004-139B-28	Sequence 28, Appl	c 922	13	1.1	253	2	US-08-520-678A-25	Sequence 25, Appl
c 850	13	1.1	40	2	US-08-811-492-28	Sequence 28, Appl	c 923	13	1.1	257	2	US-08-520-678A-24	Sequence 24, Appl
c 851	13	1.1	40	2	US-08-811-492-28	Sequence 9, Appl	c 924	13	1.1	260	2	US-08-520-678A-29	Sequence 29, Appl
c 852	13	1.1	40	5	PCT-US96-10545A-28	Sequence 28, Appl	c 925	13	1.1	266	1	US-08-594-031-32	Sequence 32, Appl
c 853	13	1.1	41	1	US-08-103-739B-5	Sequence 5, Appl	c 926	13	1.1	266	1	US-08-594-031-124	Sequence 124, App
c 854	13	1.1	41	1	US-08-474-404-5	Sequence 5, Appl	c 927	13	1.1	270	2	US-08-520-678A-30	Sequence 30, Appl
c 855	13	1.1	41	2	US-08-485-845-5	Sequence 5, Appl	c 928	13	1.1	271	2	US-08-731-272A-29	Sequence 29, Appl
c 856	13	1.1	41	2	US-08-482-714-5	Sequence 5, Appl	c 929	13	1.1	276	1	US-08-248-474-79	Sequence 79, Appl
c 857	13	1.1	41	2	US-08-482-714-5	Sequence 31, Appl	c 930	13	1.1	276	1	US-08-248-474-79	Sequence 79, Appl
c 858	13	1.1	42	1	US-08-391-000-31	Sequence 31, Appl	c 931	13	1.1	278	3	US-08-756-849-106	Sequence 106, App
c 859	13	1.1	42	1	US-08-465-980-6	Sequence 6, Appl	c 932	13	1.1	278	3	US-08-756-849-106	Sequence 106, App
c 860	13	1.1	42	1	US-08-592-406-3	Sequence 3, Appl	c 933	13	1.1	285	2	US-08-520-678A-20	Sequence 20, Appl
c 861	13	1.1	42	2	US-08-741-931-31	Sequence 31, Appl	c 934	13	1.1	291	4	US-09-172-711-39	Sequence 39, Appl
c 862	13	1.1	42	2	US-09-053-303-6	Sequence 6, Appl	c 935	13	1.1	298	2	US-08-520-678A-21	Sequence 21, Appl
c 863	13	1.1	42	5	PCT-US95-07093-6	Sequence 6, Appl	c 936	13	1.1	300	1	US-08-727-034-4	Sequence 4, Appl
c 864	13	1.1	46	2	US-08-860-882A-9	Sequence 9, Appl	c 937	13	1.1	300	1	US-08-242-098-10	Sequence 10, Appl
c 865	13	1.1	48	3	US-08-855-058-18	Sequence 18, Appl	c 938	13	1.1	300	2	US-08-637-759B-28	Sequence 28, Appl
c 866	13	1.1	51	3	US-08-855-058-15	Sequence 15, Appl	c 939	13	1.1	305	3	US-08-871-355A-28	Sequence 13, Appl
c 867	13	1.1	51	3	US-08-855-058-17	Sequence 17, Appl	c 940	13	1.1	305	1	US-08-207-904-13	Sequence 1, Appl
c 868	13	1.1	51	3	US-08-855-058-16	Sequence 16, Appl	c 941	13	1.1	315	4	US-08-945-424-1	Sequence 1, Appl
c 869	13	1.1	61	2	US-08-486-953A-50	Sequence 50, Appl	c 942	13	1.1	321	2	US-08-520-678A-23	Sequence 23, Appl
c 870	13	1.1	61	4	US-09-046-247-50	Sequence 50, Appl	c 943	13	1.1	339	2	US-08-892-880-11	Sequence 11, Appl
c 871	13	1.1	66	2	US-08-788-943A-16	Sequence 16, Appl	c 944	13	1.1	356	4	US-09-060-756-688	Sequence 22, Appl
c 872	13	1.1	67	1	US-08-374-641-25	Sequence 25, Appl	c 945	13	1.1	357	2	US-08-294-143-1	Sequence 1, Appl
c 873	13	1.1	75	1	US-08-153-799-9	Sequence 9, Appl	c 946	13	1.1	357	2	US-09-256-331-1	Sequence 1, Appl
c 874	13	1.1	75	3	US-08-416-711-6	Sequence 6, Appl	c 947	13	1.1	366	1	US-08-040-204-7	Sequence 7, Appl
c 875	13	1.1	75	3	US-08-755-587-3	Sequence 3, Appl	c 948	13	1.1	366	1	US-08-853-659A-11	Sequence 11, Appl
c 876	13	1.1	81	5	PCT-US92-06821A-13	Sequence 13, Appl	c 949	13	1.1	374	1	US-08-592-406-15	Sequence 15, Appl
c 877	13	1.1	82	1	US-07-744-282C-13	Sequence 13, Appl	c 950	13	1.1	387	4	US-09-474-922A-2	Sequence 2, Appl
c 878	13	1.1	82	4	US-09-014-416-55	Sequence 55, Appl	c 951	13	1.1	403	1	US-08-408-133-4	Sequence 4, Appl
c 879	13	1.1	90	1	US-08-123-702-16	Sequence 16, Appl	c 952	13	1.1	403	1	US-08-454-683-4	Sequence 4, Appl
c 880	13	1.1	92	3	US-09-129-740-7	Sequence 7, Appl	c 953	13	1.1	403	1	US-08-116-778E-20	Sequence 20, Appl
c 881	13	1.1	98	2	US-08-731-272A-1	Sequence 1, Appl	c 954	13	1.1	403	2	US-08-438-562-20	Sequence 20, Appl
c 882	13	1.1	98	2	US-08-731-272A-2	Sequence 2, Appl	c 955	13	1.1	403	2	US-08-483-528B-20	Sequence 20, Appl
c 883	13	1.1	98	2	US-08-731-272A-3	Sequence 3, Appl	c 956	13	1.1	442	2	US-08-673-799C-20	Sequence 20, Appl
c 884	13	1.1	98	2	US-08-731-272A-4	Sequence 4, Appl	c 957	13	1.1	450	3	US-09-040-984-24	Sequence 24, Appl
c 885	13	1.1	98	2	US-08-731-272A-5	Sequence 5, Appl	c 958	13	1.1	450	3	US-09-118-841B-1	Sequence 1, Appl
c 886	13	1.1	101	2	US-08-520-678A-1	Sequence 1, Appl	c 959	13	1.1	453	2	US-09-414-439-1	Sequence 1, Appl
c 887	13	1.1	101	2	US-08-520-678A-2	Sequence 2, Appl	c 960	13	1.1	457	1	US-08-503-800A-2	Sequence 2, Appl
c 888	13	1.1	101	2	US-08-520-678A-3	Sequence 3, Appl	c 961	13	1.1	457	2	US-08-518-878B-34	Sequence 34, Appl
c 889	13	1.1	101	2	US-08-520-678A-4	Sequence 4, Appl	c 962	13	1.1	457	2	US-08-807-861A-34	Sequence 34, Appl
c 890	13	1.1	101	2	US-08-520-678A-5	Sequence 5, Appl	c 963	13	1.1	457	3	US-08-470-868A-34	Sequence 34, Appl
c 891	13	1.1	101	4	US-09-014-416-58	Sequence 1, Appl	c 964	13	1.1	458	3	US-09-210-681-34	Sequence 34, Appl
c 892	13	1.1	105	4	US-09-014-416-62	Sequence 4, Appl	c 965	13	1.1	458	3	US-08-946-719A-34	Sequence 34, Appl
c 893	13	1.1	119	2	US-08-731-272A-27	Sequence 27, Appl	c 966	13	1.1	471	2	US-08-742-023-8	Sequence 8, Appl
c 894	13	1.1	124	4	US-07-594-921C-11	Sequence 11, Appl	c 967	13	1.1	471	2	US-08-968-503-8	Sequence 11, Appl
c 895	13	1.1	124	4	US-08-307-401-1	Sequence 1, Appl	c 968	13	1.1	471	2	US-08-465-380-11	Sequence 11, Appl
c 896	13	1.1	138	1	US-08-460-194-1	Sequence 1, Appl	c 969	13	1.1	471	2	US-08-480-478-40	Sequence 40, Appl
c 897	13	1.1	138	5	PCT-US93-11783-1	Sequence 1, Appl	c 970	13	1.1	471	2	US-08-486-397-11	Sequence 11, Appl
c 898	13	1.1	144	6	5198542-10	Sequence 65, Appl	c 971	13	1.1	471	2	US-08-486-399-11	Sequence 11, Appl
c 899	13	1.1	167	6	5198542-10	Sequence 1, Appl	c 972	13	1.1	471	2	US-08-461-965-11	Sequence 11, Appl
c 900	13	1.1	168	5	PCT-US94-09653A-26	Sequence 26, Appl	c 973	13	1.1	471	2	US-08-326-110A-40	Sequence 40, Appl
c 901	13	1.1	171	5	PCT-US94-09653A-31	Sequence 31, Appl	c 974	13	1.1	471	2		
c 902	13	1.1	183	1	US-08-322-962-2	Sequence 2, Appl	c 975	13	1.1	471	2		
c 903	13	1.1					c 976	13	1.1	471	2		

ALIGNMENTS

FEATURE:	NAME/KEY:	CDS
LOCUS	LOCATION:	1..741
GENES	ORF-08-937-271-3	
Query Match	34.7%;	Score 402; DB 3; Length 741;
Best Local Similarity	100.0%;	Pred. No. 5.1e-195;
Matches 402;	Conservative	0; Mismatches 0; Indels 0; Gaps
QY	190	ggatcgcgcgtgactagggggtacaataaattgacccgcgaagaagcaaaagaagactcttgac 249
DB	340	GGATTCGCCCGTGACTAGGGGTACAAATAAATGACCGCCAAAGACCAAAAGAACTCTTGAC 399
QY	250	aagtatagctagaagaacctgacttaaaactaagaatgaagggttaaaaactgagaat 309
DB	400	ANGTATGAGCTGAAACCATGCTTAAAACTAAGAATGAAGGGTTAAAAACTTGAGAAT 459
QY	310	gaagggttaaaaactdgagaatgaagggttaaaaactgagaatgaagggttaaaaactgaag 369
DB	460	GAAGGGTTAAAAACTTGAGAAATGAAGGGTTAAAAACTTGAGAAATGAAGGGTTAAAAACTGAG 519
QY	370	gtgcacagagtggtttctctaggggacggtagaacaacccgcacaaagcacgagaactcttt 429
DB	520	GTTCGACAGAGTGTTTCTCTAGCGGACGGTTAGAAAACCCGGACAAAGCACGAGAACTTCTT 579
QY	430	aacaagtagtcagctagagaactctatgttcacaagctaatatgacaagttaacatgaga 489
DB	580	ACAAGTATGAGTGAAGAACTCTATGTGTACAAAGCTAATAATGACAAGTTACCATGGAGA 639
QY	490	gtgcgtatactagcctacgcagaagaatgaagctaaaaaaaattatgtacgatctttgac 549
DB	640	GTGCGGTTATACTAGGCATACGCCAGAGAATAGCTTAAAAAAATTAITGACGATCTTGAC 699
QY	550	gcaaaagaacatgaattacaacaacagatgagaagttatct 591
DB	700	GTAAAGAACATGTAATTACAAACACGAATGACGAAGTTATCT 741

```

1  RESULT 2
2  US-08-937-271-1
3  ; Sequence 1, Application US/08937271
4  ; Patent NO. 6063386
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Dale, James B.
7  ; APPLICANT: Lederer, James W.
8  ; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
9  ; TITLE OF INVENTION: VACCINE
10 ; NUMBER OF SEQUENCES: 40
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: SEED AND BERRY
13 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
14 ; CITY: Seattle
15 ; STATE: Washington
16 ; COUNTRY: USA
17 ; ZIP: 98104
18 ; COMPUTER READABLE FORM:
19 ; MEDIUM TYPE: Floppy disk
20 ; COMPUTER: IBM PC compatible
21 ; OPERATING SYSTEM: PC-DOS/MS-DOS
22 ; SOFTWARE: Patentin Release #1.0, Version #1.30
23 ; CURRENT APPLICATION DATA:
24 ; APPLICATION NUMBER: US/08/937,271
25 ; FILING DATE: 15-SEP-1997
26 ; CLASSIFICATION: 424
27 ; ATTORNEY/AGENT INFORMATION:
28 ; NAME: Rosenman, Stephen J.
29 ; REGISTRATION NUMBER: 43, 058
30 ; REFERENCE/DOCKET NUMBER: 481112.405C1
31 ; TELECOMMUNICATION INFORMATION:
32 ; TELEPHONE: (206) 622-4900
33 ; TELEFAX: (206) 682-6031
34 ; INFORMATION FOR SEQ ID NO: 1:

```

```

:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 861 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pyogenes
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..861
: US-08-937-271-1

```

Query Match	20.7%	Score 240;	DB 3;	Length 861;
Best Local Similarity	100.0%;	Pred. No. 9,5e-113;		
Matches 240;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 910	gtcgcgactagctctgacagacatactctggaaaaagtcacaaagacgtgctgacaaagttt	969		
Db 4	gtcgcgactagctctgacagacatactctctggaaaaagtcacaaagacgtgctgacaaagttt	63		
QY 970	gagatagaacaaatacgtttaaacttaagaaatagtgacttaagttttaataataaaqcg	1029		
Db 64	gagatagaacaaatacgtttaaacttaagaaatagtgacttaagttttaataataaaqcg	123		
QY 1030	ttaaaagatcataatgatgatgttaactcgaagagttagttaaagcttaaaagagaactacgt	1089		
Db 124	ttaaaagatcataatgatgatgtttaaactcgaagagttagttaaagcttaaaagagaactacgt	183		
QY 1090	aaaaatgataaaatcactatctgaaaagacctaataaattcagaataattagagcgacgtaag	1149		
Db 184	aaaaatgataaaatcactatctgaaaagacctaataaattcagaataattagagcgacgtaag	243		

```

RESULT      3
US-08-937-271-14
; Sequence 14, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 base pairs

```

```

: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pyogenes
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 1..519
US-08-937-271-14

```

Query Match	18.7%;	Score 217;	DB 3;	Length 522;
Best Local Similarity	100.0%;	Pred. No. 4.6e+101;		
Matches 217; Conservative	0;	Mismatches	0;	Indels
			0;	Gaps
QY 910	gtcgcgacttagtgctctcagacagatactctgaaaagaagtacaaagcgtgctgacaagttt	969		
Db				
4	GTGCGGACTAGGCTCTCAGACAGATACTCTGAAAAGAAGTACAAAGCGTGTCACAAGTTT	63		
QY 970	gaagatagaaaaacaatctcgttaaaacttaagaatagtgacttaagtttttaataataaagcg	1029		
Db				
64	GAGATAGAAAAACAATACGTTAAAACTTTAAGAATAGTGACTTAAGTTTTAAATAAAGCG	123		
QY 1030	ttaaaagatcaataatgatgagtgtaactgaagagcttgagtaagtctaagagaagaactacgt	1089		
Db				
124	TTAAAAGATCATAAATGATGAGTTTAACTTGAAGAGGTTTGAGTAATGCTTAAGAGAAACTAGCT	183		
QY 1090	aaaaatgataaatcaactatctgaaaaagcgtagtaaaa	1126		
Db				
184	AAAAATGATAATCACTATCTGAAAAAGCTAGTAAAA	220		

RESULT 4
US-08-937-271-12
; Sequence 12, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..822
;
; US-08-937-271-5
;
;
; Query Match 18.7%; Score 217; DB 3; Length 822;
; Best Local Similarity 100.0%; Pred. No. 4.6e-101;
; Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 910 gtccgactagcttcacagactactctgaaaaagtacaaagcgtgctgacaaagttt 969
; DB 4 TTCGCCACATAGGTCACGACACATACCTGCAAAAGCTACAAAGTTC 63
;
; QY 970 gagatagaaaaacaatacgtttaaaacttaagaatagtgacttaagttttaataataaagcg 1029
; DB 64 GAGATAGAAAAACAATACGTTAAACCTTAAGAATAGTGACTTTAAGTTTAAATAATAAAGCG 123
;
; QY 1030 ttaaagatcataatgatagtgtaactgaagagtgtagtaactgtaagaagaaactacgt 1089
; DB 124 TTAAGAATCATATAGTATGAGTTAACTGAGAGTTGAGTAATGCTAAAGAGAAACTACGT 183
;
; QY 1090 aaaaatgataatcacatctatctgaaaaagctagtaaaa 1126
; DB 184 AAAAATGATAAATCAGTATCTGAAAAAGCTAGTAAAA 220
;
;
; RESULT 6
; US-08-937-271-16
; Sequence 16, Application US/08937271
; Patent No. 6063386
;
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Avenue, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;

```

```

;
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
;
US-08-937-271-12

Query Match 18.7%; Score 217; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 4.6e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtccgactaggcttcagacagatactctgaaaaagatacaagaagctgctgacaagttt 969
Db 406 GTCGCGACTAGGTCCTCAGACAGATACTCTGGAANAAGTACAAAGCGTGCTGCAAGTTT 465

QY 970 gagatagaaaaacaatacgttaaaacttaagaatagtagtactaaagtttttaataaaagcg 1029
Db 466 GAGATAGAAAACAATAACGTTAAACACTTAAAGATAGTGACTTAAGTTTAAATAAAGCG 525

QY 1030 ttaaaagatcataatgatgagtttaactgaagagtgagtaagtgaatgaagaacactacgt 1089
Db 526 TTAAGACATCAATGATGATGAGTTAACTGAAGAGTTGAGTAATGCTAAAGAGAACTACGT 585

QY 1090 aaaaatgataaatcacatctatctgaaaaagctagtaaaa 1126
Db 586 AAAAATGATAATCACTATCTGAAAAGCTAGTAAAA 622

RESULT 5
US-08-937-271-5
; Sequence 5, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SERD and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)

```

ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1029
US-08-937-271-16

Query Match 18.7%; Score 217; DB 3; Length 1029;
Best Local Similarity 100.0%; Pred. No. 4.5e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 910 gtccgactaggtctcagacagatcttggaataaaatcagaacgtctgacaagttt 969
Db 406 GTCGCGACTAGTCTCAGACAGATCTCTGCAAAAGTACAAGAACCTGCTGACAAGTTT 465
QY 970 gagatagaaaaacatacgtttaaaacttaagaatagtgacttaagtttaataataaagcg 1029
Db 466 GAGATAGAAAACAATACGTTAAAACCTTAAGANTAGTGACTTAAGTTTATAATAAAGCG 525
QY 1030 ttaaaagatcataatgatgttaagtttaactgaagagtttgagtaatgctaaagagaaactacgt 1089
Db 526 TTAAGAGATCAATATGATGATTAACCTGAAGAGTTGAGTAATGCTAAAGAGAACTACCT 585
QY 1090 aaaaatgataaatacactatctgaaaaagctagtataaa 1126
Db 586 AAAAATGATAATCACTATCTGAAAAAGCTAGTAAAA 622

RESULT 7
US-08-795-475-5
; Sequence 5, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795.475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1329
US-08-795-475-5

Query Match 13.0%; Score 151; DB 2; Length 1332;
Best Local Similarity 100.0%; Pred. No. 1.5e-67;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacggatggttaactcctaggaagttatagaagatcttcagcaacaatcccgaata 657
Db 1 AACGGTGATGTAATCTCTAGGGAAGTTTATAGAAGATCTTCAGCAACAATCCCGCAATA 60
QY 658 caaataatacgtttacgtcacgaaacaagactttaaagcgagattagaatgcaatg 717
Db 61 CAATATATACGTTTACGTCAGCAAAACAAGGACTTAAAGCGAGATTAGAGATGCAATG 120
QY 718 gaattgcaggaagagattttaagagagctg 748
Db 121 GAAGTTGCAGAGAGAGATTTTAAAGAGAGCTG 151

RESULT 8
US-08-937-271-7
; Sequence 7, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937.271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
US-08-937-271-7

Query Match 4.6%; Score 53; DB 3; Length 561;
Best Local Similarity 100.0%; Pred. No. 9e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 ggaaccgctgactaggggtacataataatgaccgcgaagcaaaagcaagcaag 242
Db 139 GGATCCCGCGTACTAGGGGTACATAATAATGACCCGCAAGAGCAAAAGAGC 191

RESULT 9
US-08-937-271-19
; Sequence 19, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..201
; US-08-937-271-19

Query Match 4.5%; Score 52; DB 3; Length 204;
Best Local Similarity 100.0%; Pred. No. 3e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 ggaaccgctgactaggggtacataataatgaccgcgaagcaaaagcaag 241
Db 49 GGATCCCGCGTACTAGGGGTACATAATAATGACCCGCAAGAGCAAAAGAGC 100

RESULT 10
US-08-937-271-21
; Sequence 21, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..405
US-08-937-271-21

Query Match 4.5%; Score 52; DB 3; Length 408;
Best Local Similarity 100.0%; Pred. No. 2.9e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 ggaaccgctgactaggggtacataataatgaccgcgaagcaaaagcaag 241
Db 49 GGATCCCGCGTACTAGGGGTACATAATAATGACCCGCAAGAGCAAAAGAGC 100

RESULT 11
US-08-937-271-9
; Sequence 9, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 918 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..915
US-08-937-271-9

Query Match 4.5%; Score 52; DB 3; Length 918;
Best Local Similarity 100.0%; Pred. No. 2.8e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 190 ggatccgcctgactaggggtacataaataagccgcgaagcaagaagaag 241
|||||
Db 49 GGATCCGCGTGACTAGGGGTACATAAATGACCCGCAAGACCAAGAAAG 100

RESULT 12
US-08-937-271-26
Sequence 26, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-937-271-26

Query Match 3.3%; Score 38; DB 3; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.1e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 gatccgcctgactaggggtacataaataagccgcgaacaa 228
|||||
Db 1 GATCCGCGTGACTAGGGGTACATAAATGACCCGCAA 38

RESULT 13
US-08-937-271-28
Sequence 28, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-937-271-28

Query Match 2.9%; Score 34; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 371 tcgacacagagtcttcttaggggacgtagaataa 404
|||||
Db 1 TCGACAGAGTGTTCCTAGGGGACGCTAGAAAA 34

RESULT 14
US-08-937-271-38
Sequence 38, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-937-271-34

Query Match 2.8%; Score 32; DB 3; Length 54;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 491 tgcgttatactaggcaccgagaagataag 522
|||||
DB 5 TGC GTTACTAGGCATACGCCAGAGATAAG 36

Search completed: May 13, 2001, 06:16:14
Job time: 4334 sec

TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 54 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-937-271-38

Query Match 2.8%; Score 32; DB 3; Length 54;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 491 tgcgttatactaggcaccgagaagataag 522
|||||
DB 5 TGC GTTACTAGGCATACGCCAGAGATAAG 36

RESULT 15
US-08-937-271-34
Sequence 34, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:55:40 ; Search time 1035.91 Seconds
(without alignments)
9765.715 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacgatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 9623517 seqs, 4368049070 residues

Word size : 10
Total number of hits satisfying chosen parameters: 6063379

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est33.*
26: gb_est34.*
27: gb_est35.*
28: gb_est36.*
29: gb_est37.*
30: gb_est38.*
31: gb_est39.*
32: gb_est40.*
33: em_estba.*
34: em_estfun.*
35: em_esthum1.*
36: em_esthum2.*
37: em_esthum3.*
38: em_esthum4.*
39: em_esthum5.*
40: em_esthum6.*
41: em_esthum7.*
42: em_esthum8.*
43: em_esthum9.*
44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estini.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estrol.*
83: em_estrol2.*
84: em_estrol3.*
85: em_estrol4.*
86: em_estrol5.*
87: em_estrol6.*
88: em_estrol7.*
89: em_estrol8.*
90: em_estrol9.*
91: em_estrol10.*
92: em_estrol11.*
93: em_estrol12.*
94: em_estrol13.*
95: em_estrol14.*
96: em_estrol15.*
97: em_estrol16.*
98: em_estrol17.*
99: em_estrol18.*
100: em_estrol19.*
101: em_estrol20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est41.*
111: gb_est42.*
112: gb_est43.*
113: gb_est44.*
114: gb_est45.*
115: gb_est46.*
116: gb_est47.*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
```

```
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	1.7	188	104	AI971561
C 2	20	1.7	453	112	AW229282
C 3	20	1.7	458	203	AO165852
C 4	20	1.7	465	104	AI970049
C 5	20	1.7	604	1	AA013829
C 6	20	1.7	608	165	BE311177
C 7	20	1.7	636	18	AI322463
C 8	20	1.7	665	13	AA880933
C 9	20	1.7	842	226	AZ693348
C 10	20	1.7	893	225	AZ675337
C 11	19	1.6	175	8	AA492011
C 12	19	1.6	272	147	BF362820
C 13	19	1.6	291	109	AV136611
C 14	19	1.6	331	152	HI8381
C 15	19	1.6	365	23	AI638554
C 16	19	1.6	378	169	BF832653
C 17	19	1.6	415	136	BE576549
C 18	19	1.6	419	206	AQ401806
					AI971561 wq87d08.x
					AW229282 uo08g01.y
					AQ165852 HS_3070.A
					AI970049 wq80d04.x
					AA013829 mh04h05.r
					BE311177 601057718
					AI322463 mh04h05.y
					AA880933 vx34805.r
					AZ693348 ENTMLJ05FF
					AZ675337 ENT1W55TR
					AA492011 ne72h01.s
					BF362820 CM4-NN007
					AV136611 AV136611
					HI8381 yn49q01.r1
					AI638554 ts50a01.x
					BF832653 PM0-WT091
					BE576549 Gc41607.y
					AQ401806 HS_5045.A

165	17	1.5	243	107	AU096015	AU096015	AU096015	238	17	1.5	429	219	AZ294965	AZ294965	RPCI-23-1
c 166	17	1.5	246	218	AZ258517	RPCI-23-1	AZ258517	c 239	17	1.5	430	211	AQ759121	AQ759121	HS_2249_B
c 167	17	1.5	265	121	AW901883	QV0-NN102	AW901883	c 240	17	1.5	431	216	AQ140062	AQ140062	SP_0006_A
c 168	17	1.5	275	161	BB603662	BB603662	BB603662	c 241	17	1.5	433	119	AW735219	AW735219	pa49b04.y
c 169	17	1.5	278	21	AI522595	fb60a06.x	AI522595	242	17	1.5	435	112	AW185289	AW185289	se89e07.y
c 170	17	1.5	282	121	AW901853	QV0-NN102	AW901853	243	17	1.5	436	147	BF395386	BF395386	UI-R-CMO-
c 171	17	1.5	282	224	AZ609199	IM0433A20	AZ609199	c 244	17	1.5	436	201	AQ003209	AQ003209	RPCI11-30
c 172	17	1.5	283	122	BB000878	BB000878	BB000878	c 245	17	1.5	437	7	AA435910	AA435910	zt87b09.s
c 173	17	1.5	287	109	AV090112	AV090112	AV090112	c 246	17	1.5	437	14	AI008196	AI008196	EST202647
c 174	17	1.5	296	130	BB330254	BB330254	BB330254	c 247	17	1.5	437	172	BG044199	BG044199	aaa25b04.
c 175	17	1.5	300	17	AI241270	qk16b12.x	AI241270	248	17	1.5	439	142	BE988600	BE988600	UI-M-CGOp
c 176	17	1.5	300	21	AI490518	EST249052	AI490518	c 249	17	1.5	440	122	AW991137	AW991137	SSS0339.S
c 177	17	1.5	300	27	AV302376	AV302376	AV302376	c 250	17	1.5	442	208	AQ604937	AQ604937	HS_21335-B
c 178	17	1.5	304	128	BB255576	BB255576	BB255576	c 251	17	1.5	446	201	AQ062695	AQ062695	CIT-HSP-2
c 179	17	1.5	309	10	AA661238	MB356V4E0	AA661238	c 252	17	1.5	448	217	AZ151868	AZ151868	SP_0006_B
c 180	17	1.5	313	16	AI145969	AI145969	AI145969	c 253	17	1.5	448	217	AZ151868	AZ151868	SP_0112_A
c 181	17	1.5	325	154	AI214898	qg46g02.x	AI214898	c 254	17	1.5	449	158	W76615	W76615	z667c09.r1
c 182	17	1.5	330	204	AQ260536	AQ260536	AQ260536	c 255	17	1.5	450	16	AI103063	AI103063	EST212352
c 183	17	1.5	332	122	AW947375	CM3-ET004	AW947375	c 256	17	1.5	453	227	B28457	B28457	F24320FF IG
c 184	17	1.5	336	130	BB332667	BB332667	BB332667	c 257	17	1.5	455	32	AV720677	AV720677	AV720677
c 185	17	1.5	341	162	BE096357	BE096357	BE096357	c 258	17	1.5	456	32	AV718700	AV718700	AV718700
c 186	17	1.5	349	224	AZ624560	AZ624560	AZ624560	c 259	17	1.5	458	32	AV719165	AV719165	AV719165
c 187	17	1.5	351	152	D46725	D46725	D46725	c 260	17	1.5	458	32	AV719838	AV719838	AV719838
c 188	17	1.5	352	154	N59060	N59060	N59060	c 261	17	1.5	458	32	AV720880	AV720880	AV720880
c 189	17	1.5	353	16	AI108528	AI108528	AI108528	c 262	17	1.5	460	113	AW277901	AW277901	sf88b10.y
c 190	17	1.5	353	31	AV698896	AV698896	AV698896	c 263	17	1.5	462	14	AA996620	AA996620	UI-R-CO-h
c 191	17	1.5	354	17	AI208123	AI208123	AI208123	c 264	17	1.5	462	21	AI536536	AI536536	vp09a03.y
c 192	17	1.5	354	22	AI554990	AI554990	AI554990	c 265	17	1.5	463	103	AI896562	AI896562	EST266005
c 193	17	1.5	354	22	AI554990	AI554990	AI554990	c 266	17	1.5	463	201	AQ078479	AQ078479	CIT-HSP-2
c 194	17	1.5	356	152	D46300	D46300	D46300	c 267	17	1.5	464	103	AI884041	AI884041	fc72f06.x
c 195	17	1.5	357	229	CNS00XHQ	CNS00XHQ	CNS00XHQ	c 268	17	1.5	464	202	AQ144790	AQ144790	HS_3092_A
c 196	17	1.5	358	159	W96751	W96751	W96751	c 269	17	1.5	467	113	AW313779	AW313779	9309.MARC
c 197	17	1.5	360	201	AQ063981	AQ063981	AQ063981	c 270	17	1.5	467	208	AQ581981	AQ581981	RPCI-11-4
c 198	17	1.5	364	22	AI555663	AI555663	AI555663	c 271	17	1.5	468	170	BF905001	BF905001	RPCI-11-4
c 199	17	1.5	369	225	AZ647798	AZ647798	AZ647798	c 272	17	1.5	472	20	AI478942	AI478942	tm25f12.x
c 200	17	1.5	372	1	AA058059	AA058059	AA058059	c 273	17	1.5	474	32	AV718915	AV718915	AV718915
c 201	17	1.5	378	114	AW353040	AW353040	AW353040	c 274	17	1.5	477	210	AQ682731	AQ682731	HS_5531_B
c 202	17	1.5	379	218	AZ270934	AZ270934	AZ270934	c 275	17	1.5	477	224	AZ590055	AZ590055	IM0399122
c 203	17	1.5	380	216	AZ143454	AZ143454	AZ143454	c 276	17	1.5	478	17	AI218862	AI218862	qg75h05.x
c 204	17	1.5	382	201	AQ012010	AQ012010	AQ012010	c 277	17	1.5	478	203	AQ164037	AQ164037	UI-M-BHO-
c 205	17	1.5	387	218	AZ276384	AZ276384	AZ276384	c 278	17	1.5	478	137	BE656048	BE656048	UI-M-BHO-
c 206	17	1.5	388	20	AI431612	AI431612	AI431612	c 279	17	1.5	479	120	AQ164037	AQ164037	UI-M-BHO-
c 207	17	1.5	390	10	AA636833	AA636833	AA636833	c 280	17	1.5	480	169	BF775042	BF775042	sp25d08.x
c 208	17	1.5	390	121	AW902448	AW902448	AW902448	c 281	17	1.5	483	169	BF775042	BF775042	sp25d08.x
c 209	17	1.5	392	29	AV402096	AV402096	AV402096	c 282	17	1.5	483	216	BF838654	BF838654	RC3-HT097
c 210	17	1.5	392	120	AW852340	AW852340	AW852340	c 283	17	1.5	484	103	AI898165	AI898165	SP_0006_A
c 211	17	1.5	394	229	AA973344	AA973344	AA973344	c 284	17	1.5	484	203	AQ245167	AQ245167	EST267608
c 212	17	1.5	396	14	AA973344	AA973344	AA973344	c 285	17	1.5	486	19	AI373657	AI373657	HS_2008_A
c 213	17	1.5	397	16	AI101086	AI101086	AI101086	c 286	17	1.5	487	209	AQ628552	AQ628552	CITBI-EI-
c 214	17	1.5	397	22	AI627688	AI627688	AI627688	c 287	17	1.5	487	224	AZ593945	AZ593945	IM0405F16
c 215	17	1.5	397	22	AI627688	AI627688	AI627688	c 288	17	1.5	488	216	AZ122271	AZ122271	RPCI-23-4
c 216	17	1.5	397	22	AI627688	AI627688	AI627688	c 289	17	1.5	488	227	B27379	B27379	T6C12TR TAM
c 217	17	1.5	397	202	AQ123542	AQ123542	AQ123542	c 290	17	1.5	489	18	AI322111	AI322111	SHO+3MCAM
c 218	17	1.5	399	113	AW278345	AW278345	AW278345	c 291	17	1.5	489	156	R80357	R80357	y196e08.r1
c 219	17	1.5	400	104	AI996099	AI996099	AI996099	c 292	17	1.5	489	208	AQ605857	AQ605857	HS_5382_A
c 220	17	1.5	400	224	AZ602244	AZ602244	AZ602244	c 293	17	1.5	490	204	AQ030428	AQ030428	HS_2053_A
c 221	17	1.5	401	11	AA765368	AA765368	AA765368	c 294	17	1.5	491	165	BE346207	BE346207	sp20h04.y
c 222	17	1.5	404	216	AZ131200	AZ131200	AZ131200	c 295	17	1.5	493	23	AI633586	AI633586	th68h05.x
c 223	17	1.5	408	7	AA417937	AA417937	AA417937	c 296	17	1.5	493	201	AQ000465	AQ000465	CIT-HSP-2
c 224	17	1.5	408	219	AZ324633	AZ324633	AZ324633	c 297	17	1.5	493	215	AQ000465	AQ000465	CIT-HSP-2
c 225	17	1.5	409	112	AW26457	AW26457	AW26457	c 298	17	1.5	495	143	BF042670	BF042670	RPCT-23-3
c 226	17	1.5	415	30	AV558081	AV558081	AV558081	c 299	17	1.5	495	164	BE224514	BE224514	kp23h05.y
c 227	17	1.5	415	104	AI982863	AI982863	AI982863	c 300	17	1.5	500	107	AU086605	AU086605	sp20h04.y
c 228	17	1.5	415	148	BF477089	BF477089	BF477089	c 301	17	1.5	503	139	BE802304	BE802304	sr28b12.y
c 229	17	1.5	416	202	AQ110253	AQ110253	AQ110253	c 302	17	1.5	505	140	BE856686	BE856686	7f66c12.x
c 230	17	1.5	420	149	BF511862	BF511862	BF511862	c 303	17	1.5	505	149	BF544530	BF544530	UI-R-BT0-
c 231	17	1.5	421	1	AA068376	AA068376	AA068376	c 304	17	1.5	505	231	CNS040QCH	CNS040QCH	Tetraodon
c 232	17	1.5	423	115	AA444456	AA444456	AA444456	c 305	17	1.5	509	219	AZ289902	AZ289902	RPCT-23-1
c 233	17	1.5	425	175	D27171	D27171	D27171	c 306	17	1.5	512	30	AV529309	AV529309	AV529309
c 234	17	1.5	426	203	AQ239645	AQ239645	AQ239645	c 307	17	1.5	512	111	AW464206	AW464206	BP230015A
c 235	17	1.5	427	217	AI182748	AI182748	AI182748	c 308	17	1.5	513	21	AI548942	AI548942	UI-R-C3-t
c 236	17	1.5	428	221	AZ442268	AZ442268	AZ442268	c 309	17	1.5	513	207	AQ520173	AQ520173	HS_3204_A
c 237	17	1.5	429	8	AA503969	AA503969	AA503969	c 310	17	1.5	516	203	AQ207993	AQ207993	HS_3163_B

C 311	17	1.5	519	147	BF406018	UI-R-CA1-	384	17	1.5	600	163	BE120615	UI-R-CA0-
C 312	17	1.5	519	147	BF410239	UI-R-CA1-	385	17	1.5	602	148	BF427573	df9h12.y
C 313	17	1.5	519	203	AQ221556	HS_3138_B	386	17	1.5	602	204	AQ259471	nbxb0023G
C 314	17	1.5	520	110	AW049753	UI-M-BH1-	C 387	17	1.5	603	222	A2497834	AZ497834
C 315	17	1.5	520	224	AZ621913		C 388	17	1.5	604	219	AZ331760	IM00334P14
C 316	17	1.5	521	30	AV528745		C 389	17	1.5	604	219	AZ331760	IM00334P14
C 317	17	1.5	522	30	AV528745		C 390	17	1.5	607	214	AQ988887	RF000553
C 318	17	1.5	522	30	AV528745		C 391	17	1.5	608	207	AQ988887	RF000553
C 319	17	1.5	522	143	BF078346		C 392	17	1.5	611	219	AQ988887	RF000553
C 320	17	1.5	523	105	AJ388951		C 393	17	1.5	611	219	AQ988887	RF000553
C 321	17	1.5	526	103	AI894643		C 394	17	1.5	616	110	AQ100729	SP95F06.y
C 322	17	1.5	526	115	AW242368		C 395	17	1.5	616	110	AQ100729	SP95F06.y
C 323	17	1.5	526	215	AZ070918		C 396	17	1.5	616	207	AQ473825	SP95F06.y
C 324	17	1.5	528	102	AI854630		C 397	17	1.5	617	214	AQ967371	SP95F06.y
C 325	17	1.5	528	162	BE098309		C 398	17	1.5	617	214	AQ967371	SP95F06.y
C 326	17	1.5	528	162	BE098309		C 399	17	1.5	619	214	AQ967077	SP95F06.y
C 327	17	1.5	528	227	B76395		C 400	17	1.5	620	16	AI114008	SP95F06.y
C 328	17	1.5	529	9	AA557322		C 401	17	1.5	620	21	AI1483516	SP95F06.y
C 329	17	1.5	530	11	AA764545		C 402	17	1.5	622	203	AQ195271	SP95F06.y
C 330	17	1.5	530	113	AA764545		C 403	17	1.5	623	116	AW519373	SP95F06.y
C 331	17	1.5	531	208	AQ569575		C 404	17	1.5	625	103	AI898494	SP95F06.y
C 332	17	1.5	531	227	B62220		C 405	17	1.5	630	214	AQ967078	SP95F06.y
C 333	17	1.5	533	202	AQ118791		C 406	17	1.5	631	121	AW859880	SP95F06.y
C 334	17	1.5	537	13	AA893234		C 407	17	1.5	631	225	AZ663915	SP95F06.y
C 335	17	1.5	537	110	AW066660		C 408	17	1.5	632	223	AZ663915	SP95F06.y
C 336	17	1.5	540	120	AW784090		C 409	17	1.5	634	117	AW587422	SP95F06.y
C 337	17	1.5	540	205	AQ358111		C 410	17	1.5	634	117	AW587422	SP95F06.y
C 338	17	1.5	540	213	AQ889644		C 411	17	1.5	636	205	AQ378104	SP95F06.y
C 339	17	1.5	541	114	AW340318		C 412	17	1.5	638	120	AW852848	SP95F06.y
C 340	17	1.5	541	147	BF409476		C 413	17	1.5	638	120	AW852848	SP95F06.y
C 341	17	1.5	542	10	AA653278		C 414	17	1.5	641	154	BF141987	SP95F06.y
C 342	17	1.5	542	16	AI145809		C 415	17	1.5	641	154	BF141987	SP95F06.y
C 343	17	1.5	543	173	BG087749		C 416	17	1.5	642	208	AQ550440	SP95F06.y
C 344	17	1.5	543	30	AV533421		C 417	17	1.5	642	208	AQ550440	SP95F06.y
C 345	17	1.5	546	32	AV736567		C 418	17	1.5	646	31	AV649277	SP95F06.y
C 346	17	1.5	546	148	BF416442		C 419	17	1.5	646	105	AL036662	SP95F06.y
C 347	17	1.5	547	21	AI486373		C 420	17	1.5	650	111	AW175310	SP95F06.y
C 348	17	1.5	547	208	AQ584531		C 421	17	1.5	655	122	AW981465	SP95F06.y
C 349	17	1.5	547	217	AZ200634		C 422	17	1.5	657	119	AW726666	SP95F06.y
C 350	17	1.5	550	168	BF744109		C 423	17	1.5	658	227	B67272	SP95F06.y
C 351	17	1.5	554	112	AW185963		C 424	17	1.5	658	227	B67272	SP95F06.y
C 352	17	1.5	555	22	AI595425		C 425	17	1.5	660	229	CNS00UVG	SP95F06.y
C 353	17	1.5	559	30	AV606309		C 426	17	1.5	662	10	AA660764	SP95F06.y
C 354	17	1.5	560	214	AQ959610		C 427	17	1.5	664	147	BF396083	SP95F06.y
C 355	17	1.5	562	147	BF410391		C 428	17	1.5	665	118	AW691975	SP95F06.y
C 356	17	1.5	563	215	AZ064499		C 429	17	1.5	667	120	AW776579	SP95F06.y
C 357	17	1.5	564	138	BE682929		C 430	17	1.5	668	164	BE213506	SP95F06.y
C 358	17	1.5	564	204	AQ246326		C 431	17	1.5	670	201	AQ052031	SP95F06.y
C 359	17	1.5	565	122	AW974016		C 432	17	1.5	681	216	AZ098707	SP95F06.y
C 360	17	1.5	565	218	AZ237073		C 433	17	1.5	682	165	BE303637	SP95F06.y
C 361	17	1.5	566	9	AA618023		C 434	17	1.5	686	162	BE109276	SP95F06.y
C 362	17	1.5	566	20	AI435590		C 435	17	1.5	688	31	AV649339	SP95F06.y
C 363	17	1.5	566	30	AV528197		C 436	17	1.5	689	230	CNS03071	SP95F06.y
C 364	17	1.5	566	207	AQ536245		C 437	17	1.5	690	226	AZ704670	SP95F06.y
C 365	17	1.5	566	225	AZ660441		C 438	17	1.5	690	226	AZ704670	SP95F06.y
C 366	17	1.5	567	138	BE682155		C 439	17	1.5	692	222	AZ461884	SP95F06.y
C 367	17	1.5	569	10	AA649047		C 440	17	1.5	695	212	AQ851636	SP95F06.y
C 368	17	1.5	570	221	AZ396106		C 441	17	1.5	699	105	AL504958	SP95F06.y
C 369	17	1.5	571	21	AI485908		C 442	17	1.5	701	111	AW132572	SP95F06.y
C 370	17	1.5	571	220	AZ383851		C 443	17	1.5	702	229	AG011313	SP95F06.y
C 371	17	1.5	573	148	BF418272		C 444	17	1.5	703	225	AQ663913	SP95F06.y
C 372	17	1.5	574	205	AQ323789		C 445	17	1.5	707	229	AG011313	SP95F06.y
C 373	17	1.5	575	227	B28599		C 446	17	1.5	711	147	BF406295	SP95F06.y
C 374	17	1.5	583	207	AQ488033		C 447	17	1.5	719	223	AZ556019	SP95F06.y
C 375	17	1.5	585	173	BG088249		C 448	17	1.5	724	212	AQ864242	SP95F06.y
C 376	17	1.5	587	221	AQ395095		C 449	17	1.5	725	31	AV702860	SP95F06.y
C 377	17	1.5	590	214	AQ967372		C 450	17	1.5	725	162	BE038349	SP95F06.y
C 378	17	1.5	592	121	AW923126		C 451	17	1.5	725	162	BE038349	SP95F06.y
C 379	17	1.5	594	21	AI484834		C 452	17	1.5	725	208	AQ584273	SP95F06.y
C 380	17	1.5	594	208	AQ585476		C 453	17	1.5	733	231	CNS03GAG	SP95F06.y
C 381	17	1.5	595	202	AI127253		C 454	17	1.5	736	233	PT020A13U	SP95F06.y
C 382	17	1.5	595	216	AZ12821		C 455	17	1.5	743	218	AZ278839	SP95F06.y
C 383	17	1.5	598	204	AQ287568		C 456	17	1.5	748	201	AF106773	SP95F06.y

c 457	17	1.5	768	230	CNS03704	Al231781 Tetraodon	c 530	16	1.4	182	127	BB199166
c 458	17	1.5	769	217	AZ183962	AZ183962 SP_1002_A	c 531	16	1.4	185	4	AA252269
c 459	17	1.5	778	138	BE704437	BE704437 8A6_CDNA	c 532	16	1.4	187	16	AI099499
c 460	17	1.5	780	140	BE820611	BE820611 GM700012B	c 533	16	1.4	187	26	AV256348
c 461	17	1.5	782	227	B20575	B20575 T20N21-T7 T	c 534	16	1.4	188	161	BE008590
c 462	17	1.5	792	32	AV758636	AV758636 AV758636	c 535	16	1.4	191	15	AI075921
c 463	17	1.5	799	165	BE295029	BE295029 601173731	c 536	16	1.4	192	146	BF330559
c 464	17	1.5	799	213	AQ896930	AQ896930 HS_5440_A	c 537	16	1.4	192	216	AZ089880
c 465	17	1.5	802	226	AZ714791	AZ714791 RPCI-24-1	c 538	16	1.4	192	216	AZ107359
c 466	17	1.5	805	210	AQ744648	AQ744648 HS_5504_A	c 539	16	1.4	194	129	BB277358
c 467	17	1.5	815	171	BF978149	BF978149 602148076	c 540	16	1.4	194	226	AZ715945
c 468	17	1.5	822	164	BE194599	BE194599 HVSMEH008	c 541	16	1.4	197	16	AI122161
c 469	17	1.5	828	136	BE539399	BE539399 601059915	c 542	16	1.4	197	30	AV565531
c 470	17	1.5	829	19	AI355567	AI355567 ct75g03_x	c 543	16	1.4	199	133	BB445715
c 471	17	1.5	831	225	AZ680674	AZ680674 ENTHH95TF	c 544	16	1.4	200	161	BE012107
c 472	17	1.5	841	208	AQ579239	AQ579239 nbsx0084E	c 545	16	1.4	201	124	BB084553
c 473	17	1.5	845	143	BF100840	BF100840 601753775	c 546	16	1.4	205	27	AV272954
c 474	17	1.5	863	226	AZ691110	AZ691110 ENTLI86TF	c 547	16	1.4	205	131	BB361185
c 475	17	1.5	870	18	AI324413	AI324413 mm53g07_Y	c 548	16	1.4	206	9	AA598172
c 476	17	1.5	872	223	AZ530382	AZ530382 ENTCH27TF	c 549	16	1.4	207	28	AV348543
c 477	17	1.5	886	217	AZ207888	AZ207888 SP_0135_A	c 550	16	1.4	207	169	BF814314
c 478	17	1.5	889	223	AZ530064	AZ530064 ENTCTJ75TF	c 551	16	1.4	209	135	BB511497
c 479	17	1.5	892	204	AQ273902	AQ273902 nbsx0031L	c 552	16	1.4	212	218	AZ234903
c 480	17	1.5	894	225	AZ669211	AZ669211 ENTMD66TF	c 553	16	1.4	213	125	BB130907
c 481	17	1.5	897	144	BF122450	BF122450 601759744	c 554	16	1.4	217	129	BB280785
c 482	17	1.5	898	171	BF978525	BF978525 602148748	c 555	16	1.4	220	162	BE077455
c 483	17	1.5	903	223	AZ541734	AZ541734 ENTCTW40TR	c 556	16	1.4	222	171	BF990090
c 484	17	1.5	911	216	BF737150	BF737150 SP_0173_A	c 557	16	1.4	222	127	BB184792
c 485	17	1.5	913	169	BF784055	BF784055 602110005	c 558	16	1.4	222	133	BB447101
c 486	17	1.5	922	230	CNS03642	AL229547 Tetraodon	c 559	16	1.4	223	29	AV268000
c 487	17	1.5	929	169	BF786242	BF786242 602112648	c 560	16	1.4	223	108	AV067626
c 488	17	1.5	935	225	AZ669308	AZ669308 ENTMP43TF	c 561	16	1.4	223	124	BB060722
c 489	17	1.5	943	145	BF205336	BF205336 601867909	c 562	16	1.4	224	17	AI228571
c 490	17	1.5	951	226	AZ689159	AZ689159 ENTFN68TF	c 563	16	1.4	224	26	AV242300
c 491	17	1.5	955	149	BF531544	BF531544 602091790	c 564	16	1.4	225	27	AV272831
c 492	17	1.5	959	231	CNS04AYM	AL282379 Tetraodon	c 565	16	1.4	225	129	BB271994
c 493	17	1.5	963	164	BE213819	BE213819 HV_CEB000	c 566	16	1.4	226	130	BB327889
c 494	17	1.5	970	229	CNS007N7	AL067241 Drosophila	c 567	16	1.4	227	27	AV271158
c 495	17	1.5	973	225	AZ685494	AZ685494 ENTIX82TF	c 568	16	1.4	227	201	AQ012235
c 496	17	1.5	990	174	BG159086	BG159086 602320518	c 569	16	1.4	227	123	BB038980
c 497	17	1.5	993	216	AI139645	AI139645 SP_0158_A	c 570	16	1.4	229	126	BB149844
c 498	17	1.5	999	229	CNS00KYP	AL078402 Drosophila	c 571	16	1.4	229	126	BB149844
c 499	17	1.5	1016	232	CNS050M6	AL349431 Tetraodon	c 572	16	1.4	229	127	BB202259
c 500	17	1.5	1055	230	CNS02D40	AL191961 Tetraodon	c 573	16	1.4	229	164	BE198780
c 501	17	1.5	1084	231	CNS04C39	AL283950 Tetraodon	c 574	16	1.4	229	170	BF877574
c 502	17	1.5	1101	232	CNS05S24	AL351301 Tetraodon	c 575	16	1.4	232	29	AV367702
c 503	17	1.5	1212	225	AZ686825	AZ686825 ENTHB41TF	c 576	16	1.4	232	119	AW740505
c 504	17	1.5	1373	139	BE780870	BE780870 601469483	c 577	16	1.4	232	170	BF899991
c 505	17	1.5	1577	174	BG262110	BG262110 602373960	c 578	16	1.4	232	173	BB038992
c 506	17	1.5	1577	40	AA778289	AA778289 z156d04_s	c 579	16	1.4	233	126	BB148125
c 507	16	1.4	52	175	C20867	C20867 HUMGS000493	c 580	16	1.4	234	27	AV271568
c 508	16	1.4	52	221	AZ435176	AZ435176 IM0222C04	c 581	16	1.4	234	150	BF553447
c 509	16	1.4	112	226	AZ733426	AZ733426 RPCI-24-1	c 582	16	1.4	235	26	AV243014
c 510	16	1.4	117	213	AQ889823	AQ889823 HS_3091_A	c 583	16	1.4	235	26	AV254691
c 511	16	1.4	121	23	AI655694	AI655694 ct14d05_x	c 584	16	1.4	235	27	AV270870
c 512	16	1.4	125	141	BE939953	BE939953 RCI-UT003	c 585	16	1.4	235	27	AV270909
c 513	16	1.4	126	219	AZ296330	AZ296330 RPCI-23-1	c 586	16	1.4	236	124	BE077346
c 514	16	1.4	134	156	R82078	R82078 e17-2.1.2-3	c 587	16	1.4	237	27	AV271022
c 515	16	1.4	138	102	AI849572	AI849572 UI-M-AH1	c 588	16	1.4	237	126	BB150132
c 516	16	1.4	141	25	AV162525	AV162525 NSL_97_H0	c 589	16	1.4	237	148	BF442387
c 517	16	1.4	143	137	BE593127	BE593127 DRFP586E	c 590	16	1.4	237	148	BF442387
c 518	16	1.4	151	105	AL079897	AL079897 DRFP586E	c 591	16	1.4	238	164	BE244051
c 519	16	1.4	151	173	BG143108	BG143108 1893102_Y	c 592	16	1.4	238	26	AV230459
c 520	16	1.4	152	207	AQ476274	AQ476274 CITBI-E1	c 593	16	1.4	238	26	AV252407
c 521	16	1.4	157	126	BE115876	BE115876 AV115876	c 594	16	1.4	238	26	AV255070
c 522	16	1.4	168	126	BB170709	BB170709 BB170709	c 595	16	1.4	238	27	AV272822
c 523	16	1.4	168	163	BE138011	BE138011 uq47c05_Y	c 596	16	1.4	238	126	BB160011
c 524	16	1.4	169	170	BF872866	BF872866 uq47c05_Y	c 597	16	1.4	238	166	BB552737
c 525	16	1.4	172	108	AV066401	AV066401 AV066401	c 598	16	1.4	238	166	BB552737
c 526	16	1.4	172	108	AV066401	AV066401 AV066401	c 599	16	1.4	239	114	AW337050
c 527	16	1.4	172	203	AQ202116	AQ202116 RPCI11-45	c 600	16	1.4	239	146	BF331468
c 528	16	1.4	175	226	AZ720574	AZ720574 RPCI-24-1	c 601	16	1.4	240	123	BB048381
c 529	16	1.4	182	14	AA957825	AA957825 UI-R-E1-f	c 602	16	1.4	241	26	AV242603

C 603	16	1.4	241	26	AV243873	AV243873	C 676	16	1.4	266	160	BF552770	BF552770
C 604	16	1.4	241	26	AV245092	AV245092	C 677	16	1.4	266	159	BF840543	BF840543
C 605	16	1.4	241	219	AZ225756	RPCI-23-1	C 678	16	1.4	267	132	BB382592	BB382592
C 606	16	1.4	242	161	AV244896	AV244896	C 679	16	1.4	268	160	BB551006	BB551006
C 607	16	1.4	242	161	BE008578	KCS-HN015	C 680	16	1.4	268	108	AV066139	AV066139
C 608	16	1.4	242	161	BE012069	KCS-BN105	C 681	16	1.4	268	131	BB374834	BB374834
C 609	16	1.4	243	22	AE020739	UI-R-AEO-	C 682	16	1.4	268	154	N22670	N22670
C 610	16	1.4	244	26	AV245018	AV245018	C 683	16	1.4	269	27	AV267967	AV267967
C 611	16	1.4	244	26	AV243742	AV243742	C 684	16	1.4	269	108	AV062870	AV062870
C 612	16	1.4	245	26	AV244977	AV244977	C 685	16	1.4	269	108	AV070313	AV070313
C 613	16	1.4	245	29	AV376523	AV376523	C 686	16	1.4	269	135	BB538248	BB538248
C 614	16	1.4	245	162	BE057315	sm01d12.y	C 687	16	1.4	270	25	AV154376	AV154376
C 615	16	1.4	246	129	BB277242	BB277242	C 688	16	1.4	270	26	AV237040	AV237040
C 616	16	1.4	246	133	BB450683	BB450683	C 689	16	1.4	270	30	AV554131	AV554131
C 617	16	1.4	247	16	AV246560	AV246560	C 690	16	1.4	270	125	BB132378	BB132378
C 618	16	1.4	247	27	AV292744	AV292744	C 691	16	1.4	270	135	BB359606	BB359606
C 619	16	1.4	248	151	BF639759	BF639759	C 692	16	1.4	271	6	AA352448	AA352448
C 620	16	1.4	248	167	BE472774	BE472774	C 693	16	1.4	271	108	AV025003	AV025003
C 621	16	1.4	249	26	AV229553	AV229553	C 694	16	1.4	271	108	AV062849	AV062849
C 622	16	1.4	249	26	AV233421	AV233421	C 695	16	1.4	271	119	AW746966	AW746966
C 623	16	1.4	249	27	AV272381	AV272381	C 696	16	1.4	271	127	BB177240	BB177240
C 624	16	1.4	250	13	AA871431	AA871431	C 697	16	1.4	271	128	BB216230	BB216230
C 625	16	1.4	250	15	AA083365	AA083365	C 698	16	1.4	272	130	BB307144	BB307144
C 626	16	1.4	250	27	AV271825	AV271825	C 699	16	1.4	272	135	BB539728	BB539728
C 627	16	1.4	250	27	AV276172	AV276172	C 700	16	1.4	273	28	AV349042	AV349042
C 628	16	1.4	250	28	AV321612	AV321612	C 701	16	1.4	273	133	BB428647	BB428647
C 629	16	1.4	251	27	AV272175	AV272175	C 702	16	1.4	273	134	BB473543	BB473543
C 630	16	1.4	251	28	AV342534	AV342534	C 703	16	1.4	273	157	T40322	T40322
C 631	16	1.4	252	27	AV270931	AV270931	C 704	16	1.4	273	163	BE135620	BE135620
C 632	16	1.4	252	128	BB215842	BB215842	C 705	16	1.4	274	13	AA869344	AA869344
C 633	16	1.4	253	26	AV244717	AV244717	C 706	16	1.4	274	28	AV340519	AV340519
C 634	16	1.4	254	13	AA869314	AA869314	C 707	16	1.4	274	108	AV070777	AV070777
C 635	16	1.4	255	123	BB028286	BB028286	C 708	16	1.4	274	170	BF912375	BF912375
C 636	16	1.4	255	124	BB056352	BB056352	C 709	16	1.4	275	125	BB32137	BB32137
C 637	16	1.4	256	155	R03248	R03248	C 710	16	1.4	275	132	BB126127	BB126127
C 638	16	1.4	257	26	AV230442	AV230442	C 711	16	1.4	275	135	BB383866	BB383866
C 639	16	1.4	257	27	AV268405	AV268405	C 712	16	1.4	275	133	BB433177	BB433177
C 640	16	1.4	258	13	AA871234	AA871234	C 713	16	1.4	276	133	BB428623	BB428623
C 641	16	1.4	258	19	AV375011	AV375011	C 714	16	1.4	277	25	AV220165	AV220165
C 642	16	1.4	259	13	AA869153	AA869153	C 715	16	1.4	277	28	AV331932	AV331932
C 643	16	1.4	259	13	AA871722	AA871722	C 716	16	1.4	277	28	AV331932	AV331932
C 644	16	1.4	259	157	T40166	T40166	C 717	16	1.4	277	131	BB352086	BB352086
C 645	16	1.4	259	171	BF999294	BF999294	C 718	16	1.4	277	133	BB420097	BB420097
C 646	16	1.4	260	121	AA887472	AA887472	C 719	16	1.4	277	134	BB474215	BB474215
C 647	16	1.4	260	124	BB083922	BB083922	C 720	16	1.4	278	124	BB088020	BB088020
C 648	16	1.4	260	125	BB118097	BB118097	C 721	16	1.4	278	152	D60434	D60434
C 649	16	1.4	260	135	BB539312	BB539312	C 722	16	1.4	278	160	BB553178	BB553178
C 650	16	1.4	260	209	AO657487	AO657487	C 723	16	1.4	279	108	AV062564	AV062564
C 651	16	1.4	261	19	AI382402	AI382402	C 724	16	1.4	279	127	BB186827	BB186827
C 652	16	1.4	261	27	AV245466	AV245466	C 725	16	1.4	280	30	AV521539	AV521539
C 653	16	1.4	261	26	AV245466	AV245466	C 726	16	1.4	281	108	AV071100	AV071100
C 654	16	1.4	261	27	AV271358	AV271358	C 727	16	1.4	281	115	AA481195	AA481195
C 655	16	1.4	261	27	AV271510	AV271510	C 728	16	1.4	281	124	BB065212	BB065212
C 656	16	1.4	262	26	AV246367	AV246367	C 729	16	1.4	281	135	BB511752	BB511752
C 657	16	1.4	262	27	AV271316	AV271316	C 730	16	1.4	281	135	BB529879	BB529879
C 658	16	1.4	262	102	AI854078	AI854078	C 731	16	1.4	281	147	BF378303	BF378303
C 659	16	1.4	262	108	AV049476	AV049476	C 732	16	1.4	282	108	AV065200	AV065200
C 660	16	1.4	262	168	BB223033	BB223033	C 733	16	1.4	282	124	BB074943	BB074943
C 661	16	1.4	262	223	AA576378	AA576378	C 734	16	1.4	282	124	BB083605	BB083605
C 662	16	1.4	263	27	AV271448	AV271448	C 735	16	1.4	282	128	BB216808	BB216808
C 663	16	1.4	263	109	AV124749	AV124749	C 736	16	1.4	282	128	BB230517	BB230517
C 664	16	1.4	263	122	BB001761	BB001761	C 737	16	1.4	283	124	BB078698	BB078698
C 665	16	1.4	264	108	AV066052	AV066052	C 738	16	1.4	283	126	BB159026	BB159026
C 666	16	1.4	264	108	AV066364	AV066364	C 739	16	1.4	283	130	BB308321	BB308321
C 667	16	1.4	265	28	AV328079	AV328079	C 740	16	1.4	283	202	AA099083	AA099083
C 668	16	1.4	265	107	AV009854	AV009854	C 741	16	1.4	284	110	AA063386	AA063386
C 669	16	1.4	265	131	BB356328	BB356328	C 742	16	1.4	284	120	AA835031	AA835031
C 670	16	1.4	266	26	AV243343	AV243343	C 743	16	1.4	284	129	BB271995	BB271995
C 671	16	1.4	266	27	AV266877	AV266877	C 744	16	1.4	284	216	AZ140574	AZ140574
C 672	16	1.4	266	27	AV277096	AV277096	C 745	16	1.4	284	230	CNS020WU	CNS020WU
C 673	16	1.4	266	108	AV065642	AV065642	C 746	16	1.4	285	27	AV296554	AV296554
C 674	16	1.4	266	125	BB128028	BB128028	C 747	16	1.4	285	105	AL045125	AL045125
C 675	16	1.4	266	131	BB349942	BB349942	C 748	16	1.4	285	108	AV056389	AV056389

C 749	16	1.4	285	108	AV064405	AV064405	C 822	16	1.4	298	108	AV051971	AV051971
C 750	16	1.4	285	122	BB011370	BB011370	C 823	16	1.4	298	108	AV055939	AV055939
C 751	16	1.4	285	127	BB182160	BB182160	C 824	16	1.4	298	108	AV069763	AV069763
C 752	16	1.4	285	129	BB262899	BB262899	C 825	16	1.4	298	124	BB083147	BB083147
C 753	16	1.4	285	134	BB483586	BB483586	C 826	16	1.4	298	128	BB248536	BB248536
C 754	16	1.4	285	135	BB526826	BB526826	C 827	16	1.4	298	130	BB307259	BB307259
C 755	16	1.4	285	163	BB135579	BB135579	C 828	16	1.4	299	108	AV064587	AV064587
C 756	16	1.4	286	28	AV341178	AV341178	C 829	16	1.4	299	123	BB040867	BB040867
C 757	16	1.4	286	29	AV364534	AV364534	C 830	16	1.4	299	128	BB221106	BB221106
C 758	16	1.4	286	108	AV055022	AV055022	C 831	16	1.4	299	129	BB280943	BB280943
C 759	16	1.4	286	108	AV064359	AV064359	C 832	16	1.4	300	13	AA884595	AA884595
C 760	16	1.4	286	128	BB237874	BB237874	C 833	16	1.4	300	124	BB079779	BB079779
C 761	16	1.4	286	134	BB472722	BB472722	C 834	16	1.4	300	175	C34491	C34491
C 762	16	1.4	286	135	BB530035	BB530035	C 835	16	1.4	300	175	C34491	C34491
C 763	16	1.4	286	135	BB530035	BB530035	C 836	16	1.4	300	175	C34491	C34491
C 764	16	1.4	287	108	AV064395	AV064395	C 837	16	1.4	301	108	AV069888	AV069888
C 765	16	1.4	287	124	BB067542	BB067542	C 838	16	1.4	301	108	AV069888	AV069888
C 766	16	1.4	287	124	BB067542	BB067542	C 839	16	1.4	301	108	AV069888	AV069888
C 767	16	1.4	287	156	BB079287	BB079287	C 840	16	1.4	301	129	BB282508	BB282508
C 768	16	1.4	287	166	BB079287	BB079287	C 841	16	1.4	301	129	BB282508	BB282508
C 769	16	1.4	287	170	BB079287	BB079287	C 842	16	1.4	302	131	BB346658	BB346658
C 770	16	1.4	288	108	AV064747	AV064747	C 843	16	1.4	302	131	BB346658	BB346658
C 771	16	1.4	288	122	BB013377	BB013377	C 844	16	1.4	302	168	BB773154	BB773154
C 772	16	1.4	288	131	BB566527	BB566527	C 845	16	1.4	302	168	BB773154	BB773154
C 773	16	1.4	288	135	BB566527	BB566527	C 846	16	1.4	303	9	AA614866	AA614866
C 774	16	1.4	288	137	BB566526	BB566526	C 847	16	1.4	303	14	AA969986	AA969986
C 775	16	1.4	289	108	AV061939	AV061939	C 848	16	1.4	303	125	BB125963	BB125963
C 776	16	1.4	289	108	AV061939	AV061939	C 849	16	1.4	303	128	BB219076	BB219076
C 777	16	1.4	289	127	BB190706	BB190706	C 850	16	1.4	303	135	BB512702	BB512702
C 778	16	1.4	289	129	BB281643	BB281643	C 851	16	1.4	303	232	PT011K050	PT011K050
C 779	16	1.4	289	134	BB485273	BB485273	C 852	16	1.4	304	28	AV338736	AV338736
C 780	16	1.4	289	168	BB072201	BB072201	C 853	16	1.4	304	108	AV049996	AV049996
C 781	16	1.4	290	108	AV054020	AV054020	C 854	16	1.4	304	127	BB201776	BB201776
C 782	16	1.4	290	123	BB040286	BB040286	C 855	16	1.4	304	128	BB236239	BB236239
C 783	16	1.4	290	129	BB282448	BB282448	C 856	16	1.4	304	134	BB476497	BB476497
C 784	16	1.4	291	108	AV053774	AV053774	C 857	16	1.4	305	25	AV172102	AV172102
C 785	16	1.4	291	108	AV053774	AV053774	C 858	16	1.4	305	25	AV172102	AV172102
C 786	16	1.4	291	111	BB134675	BB134675	C 859	16	1.4	305	25	AV172102	AV172102
C 787	16	1.4	291	128	BB322762	BB322762	C 860	16	1.4	306	160	BB546658	BB546658
C 788	16	1.4	292	26	AV235830	AV235830	C 861	16	1.4	306	160	BB546658	BB546658
C 789	16	1.4	292	108	AV052218	AV052218	C 862	16	1.4	306	160	BB546658	BB546658
C 790	16	1.4	292	108	AV052234	AV052234	C 863	16	1.4	307	5	AA324818	AA324818
C 791	16	1.4	292	108	AV052234	AV052234	C 864	16	1.4	307	7	AA61515	AA61515
C 792	16	1.4	292	108	AV064683	AV064683	C 865	16	1.4	307	125	BB106774	BB106774
C 793	16	1.4	292	129	BB261124	BB261124	C 866	16	1.4	307	126	BB136858	BB136858
C 794	16	1.4	292	131	BB359551	BB359551	C 867	16	1.4	307	129	BB285692	BB285692
C 795	16	1.4	292	133	BB431111	BB431111	C 868	16	1.4	307	132	BB393873	BB393873
C 796	16	1.4	292	133	BB431111	BB431111	C 869	16	1.4	308	123	BB044192	BB044192
C 797	16	1.4	292	201	AA012580	AA012580	C 870	16	1.4	308	125	BB111667	BB111667
C 798	16	1.4	292	233	AA012580	AA012580	C 871	16	1.4	308	125	BB111667	BB111667
C 799	16	1.4	293	108	AV058312	AV058312	C 872	16	1.4	308	125	BB111667	BB111667
C 800	16	1.4	293	108	AV058312	AV058312	C 873	16	1.4	308	125	BB111667	BB111667
C 801	16	1.4	293	122	BB006511	BB006511	C 874	16	1.4	309	127	BB192603	BB192603
C 802	16	1.4	293	122	BB006511	BB006511	C 875	16	1.4	309	127	BB192603	BB192603
C 803	16	1.4	293	124	BB068374	BB068374	C 876	16	1.4	310	123	BB044285	BB044285
C 804	16	1.4	294	27	AV271396	AV271396	C 877	16	1.4	311	127	BB206566	BB206566
C 805	16	1.4	294	108	AV056321	AV056321	C 878	16	1.4	311	127	BB206566	BB206566
C 806	16	1.4	294	108	AV064754	AV064754	C 879	16	1.4	311	127	BB206566	BB206566
C 807	16	1.4	294	127	BB212316	BB212316	C 880	16	1.4	312	125	BB124727	BB124727
C 808	16	1.4	294	129	BB278863	BB278863	C 881	16	1.4	312	125	BB124727	BB124727
C 809	16	1.4	294	134	BB478387	BB478387	C 882	16	1.4	312	160	BB542429	BB542429
C 810	16	1.4	295	28	AV339617	AV339617	C 883	16	1.4	313	13	AA869173	AA869173
C 811	16	1.4	295	108	AV064288	AV064288	C 884	16	1.4	313	125	BB130249	BB130249
C 812	16	1.4	295	126	BB136278	BB136278	C 885	16	1.4	313	128	BB243631	BB243631
C 813	16	1.4	295	128	BB231368	BB231368	C 886	16	1.4	313	129	BB281362	BB281362
C 814	16	1.4	296	108	AV056622	AV056622	C 887	16	1.4	314	215	A2042211	A2042211
C 815	16	1.4	296	129	BB282447	BB282447	C 888	16	1.4	314	6	AA370281	AA370281
C 816	16	1.4	296	138	BB282447	BB282447	C 889	16	1.4	314	13	AA871432	AA871432
C 817	16	1.4	296	155	BB282447	BB282447	C 890	16	1.4	314	124	BB070959	BB070959
C 818	16	1.4	296	163	BB282447	BB282447	C 891	16	1.4	314	127	BB200398	BB200398
C 819	16	1.4	296	227	BB282447	BB282447	C 892	16	1.4	314	130	BB327409	BB327409
C 820	16	1.4	297	108	AV070571	AV070571	C 893	16	1.4	315	111	AA145737	AA145737
C 821	16	1.4	297	134	BB471010	BB471010	C 894	16	1.4	315	134	BB460179	BB460179

FEATURES

Tue May 15 07:26:59 2001

```

source
1. .188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2479023"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: p773D-pac (Pharmacia) with a modified
polylinker; plasmid DNA from the normalized library
NCI_CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (cloneIDs 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT      65 a 52 c 16 g 55 t
ORIGIN

Query Match      1.7%: Score 20; DB 104; Length 188;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 gttgcaggagagatttaa 740
Db 71 GTTGCAGGAGAGATTTTAA 52

RESULT 2
AW229282/c      453 bp mRNA EST 10-DEC-1999
LOCUS      uo08g01.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2598960 5'
DEFINITION      similar to gb:M33782 TFEF PROTEIN (HUMAN);, mRNA sequence.
ACCESSION      AW229282
VERSION      AW229282.1 GI:6558578
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 453)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: uo08g01.x1
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html

MGI:1022420
Seq primer: -40RP from Gibco
High quality sequence stop: 438.
Location/Qualifiers
1. .453
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:2598960"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; site_1: Sali;

source
Query Match      1.7%: Score 20; DB 203; Length 458;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 taqtaaaattcaagaattag 1138
Db 250 TAGTAAATTCAGAAATTAG 231

RESULT 4
AI970049      465 bp mRNA EST 25-AUG-1999
LOCUS      wq80d04.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2477575 3',
DEFINITION

```

```

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      97 a 146 c 148 g 62 t
ORIGIN

Query Match      1.7%: Score 20; DB 112; Length 453;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 taggtctcacagacatactc 37
Db 435 TAGGTCTCAGACACATACTC 416

RESULT 3
AQ165852/c      458 bp DNA GSS 16-OCT-1998
LOCUS      HS_3070_A1.D10.MF CIT Approved Human Genomic Sperm Library D Homo
DEFINITION      sapiens genomic clone Plate=3070 Col=19 Row=G, DNA sequence.
ACCESSION      AQ165852
VERSION      AQ165852.1 GI:3564047
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 458)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99390589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3070 row: G column: 19
Class: BAC ends
High quality sequence stop: 458.
Location/Qualifiers
1. .458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=3070 Col=19 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      93 a 138 c 81 g 144 t 2 others
ORIGIN

Query Match      1.7%: Score 20; DB 203; Length 458;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 taqtaaaattcaagaattag 1138
Db 250 TAGTAAATTCAGAAATTAG 231

RESULT 4
AI970049      465 bp mRNA EST 25-AUG-1999
LOCUS      wq80d04.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2477575 3',
DEFINITION

```

```

mrna sequence.
ACCESSION AI970049
VERSION AI970049.1 GI:5766867
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 465)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
FEATURES
source
Location/Qualifiers
1..465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2477575"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 132 a 92 c 101 g 140 t
ORIGIN
Query Match 1.7%; Score 20; DB 104; Length 465;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 729 aagagattttaagagagctg 748
|||||
Db 7 AACAGATTTTAAAGAGCTG 26
|||||
RESULT 5
AA013829/c
LOCUS 604 bp mRNA EST 21-JAN-1997
DEFINITION mh04h05.r1 Soares mouse placenta 4NDMP13.5 14.5 Mus musculus cDNA
clone IMAGE:441561 5' similar to gb:M33782 TFEB PROTEIN (HUMAN);,
mrna sequence.
AA013829
AA013829.1 GI:1474875
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 604)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Norris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:266897
Seq primer: ETPPrimer
High quality sequence stop: 346.
FEATURES
source
Location/Qualifiers
1..604
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:441561"
/clone_lib="Soares mouse placenta 4NDMP13.5 14.5"
/sex="unknown"
/tissue_type="placenta"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTACCAATCTGAAGTGGAGCGCGCGAAATTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 134 a 187 c 166 g 117 t
ORIGIN
Query Match 1.7%; Score 20; DB 1; Length 604;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 taggtctcagacagatactc 37
|||||
Db 510 TAGGCTCTCAGACAGATAC 491
|||||
RESULT 6
BE311177/c
LOCUS 608 bp mRNA EST 26-OCT-2000
DEFINITION BE311177 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3157948 5',
mrna sequence.
BE311177
BE311177.1 GI:9174780
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 608)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM7650 row: c column: 05
 High quality sequence stop: 606.
 Location/Qualifiers

FEATURES

source

```

1. .608
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3157948"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      131 a 173 c 190 g 114 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 165; Length 608;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 18 taggtctcagacagatactc 37
|||||

```

```

DB 579 TAGGTCTCAGACAGATACTC 560

```

RESULT 7

```

AI322463/c
LOCUS      AI322463      636 bp      mRNA      EST      23-DEC-1998
DEFINITION      mh04h05.y1 Soares mouse placenta 4NBMP13.5 14.5 Mus musculus cDNA
clone IMAGE:441561 5' similar to gb:M33782 TFEB PROTEIN (HUMAN);,
mRNA sequence.
ACCESSION      AI322463      1 GI:4056892
VERSION      AI322463
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 636)
AUTHORS      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and
Waterston,R.

```

```

TITLE      The WashU-HHMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

```

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:266897

This read is a RESEQUENCE of a previously sequenced mouse clone
 correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 180.
 Location/Qualifiers

FEATURES

source

```

1. .636
/organism="Mus musculus"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:441561"
/clone_lib="Soares mouse placenta 4NBMP13.5 14.5"
/sex="unknown"
/tissue_type="placenta"
/lab_host="DH10B"
/Note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
15'.
TGTTACCAATCTGAAGTGGGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'/; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      136 a 203 c 173 g 119 t 5 others
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 18; Length 636;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 18 taggtctcagacagatactc 37
|||||

```

```

DB 512 TAGGTCTCAGACAGATACTC 493

```

RESULT 8

```

AA880933/c
LOCUS      AA880933      665 bp      mRNA      EST      26-MAR-1998
DEFINITION      vx34a05.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
IMAGE:1277072 5' similar to gb:M33782 TFEB PROTEIN (HUMAN);, mRNA
sequence.
ACCESSION      AA880933
VERSION      AA880933      1 GI:2989916
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 665)
AUTHORS      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and
Waterston,R.

```

```

TITLE      The WashU-HHMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

```

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:668872
 Seq primer: -28ml3 revl ET from Amersham
 High quality sequence stop: 238.
 Location/Qualifiers

FEATURES

source

```

1. .665
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1277072"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"

```



```

/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dt. 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-zap
XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      149 a   203 c   191 g   122 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 13; Length 665;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 taggtctcagacagatactc 37
|||||
Db 318 TAGGTCTCAGACAGATCTC 299

```

```

RESULT 9
AZ693348/c      842 bp      DNA      GSS      14-DEC-2000
LOCUS      ENTML05TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, DNA sequence.
ACCESSION      AZ693348
VERSION      AZ693348.1 GI:11830494
KEYWORDS      GSS.
SOURCE      Entamoeba histolytica.
ORGANISM      Eukaryota; Entamoebidae; Entamoeba.
REFERENCE      1 (bases 1 to 842)
AUTHORS      Loftus.B., Van Aken.S. and Fraser.C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL      HMI:IMSS sheared DNA library
COMMENT      Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 32
High quality sequence stop: 676.
Location/Qualifiers
1..842
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

```

```

BASE COUNT      300 a   82 c   77 g   383 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 226; Length 842;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 774 attatatactaaacataata 793
|||||
Db 171 ATTATATACTAAACATAATA 152

RESULT 10
AZ675337/c      893 bp      DNA      GSS      14-DEC-2000
LOCUS      ENTW55TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, DNA sequence.
ACCESSION      AZ675337
VERSION      AZ675337.1 GI:11812603
KEYWORDS      GSS.
SOURCE      Entamoeba histolytica.
ORGANISM      Eukaryota; Entamoebidae; Entamoeba.
REFERENCE      1 (bases 1 to 893)
AUTHORS      Loftus.B., Van Aken.S. and Fraser.C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL      HMI:IMSS sheared DNA library
COMMENT      Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 20
High quality sequence stop: 873.
Location/Qualifiers
1..893
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

```

```

BASE COUNT      356 a   106 c   91 g   340 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 225; Length 893;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 774 attatatactaaacataata 793
|||||
Db 801 ATTATATACTAAACATAATA 782

RESULT 11
AA492011/c      175 bp      mRNA      EST      19-AUG-1997
LOCUS

```

```

DEFINITION ne72h01.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:909841, mRNA
sequence.
ACCESSION AA492011
VERSION AA492011.1 GI:2221573
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 305 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES
source
1..175
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:909841"
/clone="NCI_CGAP_Ew1"
/tissue_type="Ewing's sarcoma"
/lab_host="DH10B"
/Note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
cDNA made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 53 a 44 c 29 g 49 t
ORIGIN
Query Match 1.6%; Score 19; DB 8; Length 175;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1086 acgtataaatgataatca 1104
|||||
Db 170 ACCTAAATGATAAATCA 152

RESULT 12
BF362820 272 bp mRNA EST 24-NOV-2000
LOCUS CM4-NN0079-050600-191-a05 NN0079 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF362820
ACCESSION BF362820.1 GI:11324845
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 272)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

```

```

JOURNAL Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics.
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4&t2=CM4-NN0079-
050600-191-a05&t3=2000-06-05&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 272.
FEATURES
source
1..272
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0079"
/dev_stage="Adult"
/Note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 105 a 38 c 39 g 90 t
ORIGIN
Query Match 1.6%; Score 19; DB 147; Length 272;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 aggaagagatttaagaga 744
|||||
Db 171 AGGAAGAGATTTTAAGAGA 189

RESULT 13
AV136611/c
LOCUS AV136611 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
DEFINITION AV136611 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810027G08, mRNA sequence.
ACCESSION AV136611
VERSION AV136611.1 GI:5322341
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 291)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomimaga,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rcl.riken.go.jp
Thermotabilization and thermoactivation of thermolabile enzymes by

```

trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

1. .291
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="2810027008"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
70 a 67 c 47 g 107 t

BASE COUNT

ORIGIN

Query Match 1.6%; Score 19; DB 109; Length 291;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 gaagggttaaaactgagg 370

|||||

Db 120 GAAGGGTTAAAACTGAGG 102

RESULT 14

HI18381/c

LOCUS

HI18381 331 bp mRNA EST 29-JUN-1995

Yn49401.r1 Soares adult brain N2b5HB55y Homo sapiens cDNA clone

IMAGE:171792 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 331)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

, R., Williamson, A., Wohlmann, P., and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 705

High quality sequence stops: 268

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 705 Std Error: 0.00

Seq primer: M13Rpl

High quality sequence stop: 268.

Location/Qualifiers

1. .331

/organism="Homo sapiens"

/db_xref="GDB:3833827"

/db_xref="taxon:9606"

/clone="IMAGE:171792"

/clone_lib="Soares adult brain N2b5HB55y"

/sex="Male"

/dev_stage="55-year old"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: brain; Vector: p77f3D (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAATCTGAAGTGGAGCGCGCTTTTITTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p77f3 vector (Pharmacia). Library went through one round of normalization to a Cot = 53. Library constructed by Bento Soares and M. Fatima Bonaldo. The adult brain RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18 hours after death which occurred in consequence of a ruptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

BASE COUNT 104 a 52 c 62 g 111 t 2 others

ORIGIN

Query Match 1.6%; Score 19; DB 152; Length 331;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1110 tgaaaaagctagtaaaatt 1128

|||||

Db 228 TGAAAAAGCTAGTAAATTT 210

RESULT 15

AI638554

LOCUS

AI638554

ts50a01.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2231976 3',

mRNA sequence.

AI638554

VERSION

AI638554.1 GI:4690788

KEYWORDS

EST.

SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 365)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2276 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 360.

Location/Qualifiers

1. .365

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2231976"

/clone_lib="NCI_CGAP_Ut1"

/tissue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

FEATURES

source

BASE COUNT 141 a 58 c 54 g 112 t
ORIGIN

Query Match 1.6%; Score 19; DB 23; Length 365;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1110 tgaataagctagtaaaatt 1128
|||||
Db 345 TGAAAAAGCTAGTAAAT 363

Search completed: May 13, 2001, 05:47:55
Job time: 3135 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 01:59:25 ; Search time 1527.71 Seconds
(without alignments)
11178.837 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pl1.*
- 13: gb_pl2.*
- 14: gb_pl3.*
- 15: gb_pl4.*
- 16: em_bal.*
- 17: em_ba2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vil2.*
- 60: gb_vil3.*
- 61: gb_vil4.*
- 62: gb_vil5.*
- 63: gb_vil6.*
- 64: gb_vil7.*
- 65: gb_vil8.*
- 66: gb_vil9.*
- 67: gb_vil10.*
- 68: gb_vil11.*
- 69: gb_vil12.*
- 70: gb_vil13.*
- 71: gb_vil14.*
- 72: gb_vil15.*
- 73: gb_vil16.*
- 74: gb_vil17.*
- 75: gb_vil18.*
- 76: gb_vil19.*
- 77: gb_vil20.*
- 78: gb_vil21.*
- 79: gb_vil22.*
- 80: gb_vil23.*
- 81: gb_vil24.*
- 82: gb_vil25.*
- 83: gb_vil26.*
- 84: gb_vil27.*
- 85: gb_vil28.*
- 86: gb_vil29.*
- 87: gb_vil30.*
- 88: gb_vil31.*
- 89: gb_vil32.*
- 90: gb_vil33.*
- 91: gb_vil34.*
- 92: gb_vil35.*
- 93: gb_vil36.*
- 94: gb_vil37.*
- 95: gb_vil38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	419	36.2	741	3	S64396
2	241.6	20.9	2131	3	STREMM24
3	175	15.1	1547	3	STRSMP3A
4	171.8	14.8	2633	3	SPU02480
5	158	13.6	1817	2	AF288752
6	158	13.6	1620	3	SPU40231
7	156.4	13.5	1465	3	STREMM3
8	154.8	13.4	1746	3	SPENM3G
9	151	13.0	229	3	S79254
10	151	13.0	237	3	SPU11940
11	151	13.0	300	3	SPU20094

12	151	13.0	1332	9	AR079002	Sequence
13	151	13.0	1494	3	SPENM1M	X62131 S.pyogenes
14	151	13.0	1708	3	SPENM1	X07933 S.pyogenes
15	149.4	12.9	251	3	SPU11997	U11997 Streptococc
16	149.4	12.9	300	3	SPU20097	U20097 Streptococc
17	149.4	12.9	300	3	SPU20099	U20099 Streptococc
18	149.4	12.9	300	3	SPU20100	U20100 Streptococc
19	147.4	12.7	907	1	AF060760	AF060760 Streptoco
20	137	11.8	694	2	AF288753	AF288753 Streptoco
21	130	11.2	249	3	SPAJ3285	AJ223285 Streptoco
22	123.8	10.7	300	3	SPU20095	U20095 Streptococc
23	123.8	10.7	1455	3	SPM11PROA	221845 S.pyogenes
24	113	9.8	261	3	SPU11945	U11945 Streptococc
25	108.8	9.4	296	3	SPU11959	U11959 Streptococc
26	106.4	9.2	451	3	SPU11986	U11986 Streptococc
27	106.4	9.2	2111	3	STRM6	M11338 Streptococc
28	105.6	9.1	301	10	I01016	I01016 Sequence 1
29	104.8	9.1	249	3	SPAJ3286	AJ223286 Streptoco
30	104.6	9.0	300	3	SPU20098	U20098 Streptococc
31	97.4	8.4	310	3	S82058	S82058 N-terminal
32	97.4	8.4	311	3	SPENMT19	X56609 S.pyogenes
33	92.8	8.0	300	3	SPU20096	U20096 Streptococc
34	92.4	8.0	549	2	AF281048	AF281048 Streptoco
35	92	7.9	664	2	AF222860	AF222860 Streptoco
36	91.6	7.9	321	3	SSGM507	X79527 Streptococc
37	91.4	7.9	1677	3	SPU39838	U39838 Streptococc
38	90.6	7.8	263	3	SPU11943	U11943 Streptococc
39	89.4	7.7	299	3	STREMM32G	L47325 Streptococc
40	85.2	7.4	225	1	AF018179	AF018179 Streptoco
41	84.6	7.3	430	2	GGU25741	U25741 Group G str
42	84.6	7.3	450	3	SSGM935	X79533 Streptococc
43	84	7.3	261	3	STREMM	L46652 Streptococc
44	82.6	7.1	7218	10	I66494	I66494 Sequence 14
45	77.6	6.7	714	2	AF325839	AF325839 Streptoco

RESULT 1

SS64396

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

emh=tetravalent M protein [Streptococcus, group A, Genomic Recombinant, 741 nt].

S64396

S64396.1 GI:408224

Streptococcus group A.

Streptococcus

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae.

1 (bases 1 to 741)

Dale, J.B., Chiang, E.Y. and Lederer, J.W.

Recombinant tetravalent group A streptococcal M protein vaccine

J. Immunol. 151 (4), 2188-2194 (1993)

93346765

GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 136070] from the original journal article. This sequence comes from Fig. 1.

FEATURES

source

gene

CDS

1..741

/organism="Streptococcus"

/db_xref="taxon:1301"

1..741

/partial

/gene="emm"

/note="tetravalent M protein"

1..741

/partial

/gene="emm"

/note="Hybrid molecule containing amino-terminal subunits of types 24, 5, 6, and 19 M proteins; This sequence comes from Fig. 1"

product="tetravalent M protein"

protein_id="AAB27745.1"

/db_xref="GI:408225"

/translation="MVATRSQDTLTLEKVOERADKFEIENNTLKKNSDLSFNKKALKDHNDELTEELSNKAEKLRKNDKSLSEKSKNQLEAKADLEKALEGAMNFSTADSAKIKTLEAKADLEGSAVTRGTINDPQRAKEALDYELENDHDKTKNEGLKTENBGLKTENEGLTENKGLKTEVDVPRPTVENDKARELLNKYDVENSMLQANNNDKLPWRVRYTRHTPEDKLIKIIDLDKAHEHQOQNEKLS"

BASE COUNT 311 a 106 c 164 g 160 t

ORIGIN

Query Match 36.2%; Score 419; DB 3; Length 741;

Best Local Similarity 78.9%; Pred. No. 1.9e-67;

Matches 585; Conservative 0; Mismatches 0; Indels 156; Gaps 1;

QY 7 atggtcggaactaggtctcaagacagatactctgaaagaaagtaacaaagctgctgacaaag 66

Db 1 ATGGTCCGGGACTAGGTCTCAGACAGACTCTCTGAAAAAGTACAAAGACCTGCTGCAAG 60

QY 67 ttgagatagaaaaacaatacgtttaaacttaagaatagtgacttaagctttaaataaaa 126

Db 61 TTTGAGATAGAAAAACAATACGTAAACTTAAGAAATAGTACTTAAGTTTAATAATAAA 120

QY 127 gcgttaagaatcataagtagagtgagtgactaaagctgagtgactaaagtgagaaacta 186

Db 121 GCGTTAAAGATCATAATGATGAGTTAACTGAAGAGTTGAGTAATGCTTAAGAGAACTA 180

QY 187 cgt----- 189

Db 181 CGTAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAAATCAAGAAATTAGAGGCAGT 240

QY 190 ----- 189

Db 241 AAGGCTGATCTTGAAAAAGCATTAGAAGCGCAATGAATTTTTCACAGCGGATCAGCT 300

QY 190 -----ggatccgcgcgtgactaggggt 210

Db 301 AAATCAAAACCTTAGAAGCAGAGAAAGCTGATCTTGAAGGATCGCGGTGACTAGGGGT 360

QY 211 acaataatgacccgcaagagagcaaaagagctcttgacaagctatgagctagaacacat 270

Db 361 ACAATAATGACCCGCAAGAGAGCAAAAGAGCTCTTTGACAAATATGAGCTAGAAAACCAT 420

QY 271 gactaaaaactaagaatgaagggttaaaaactgagaatgaagggttaaaaactgagaat 330

Db 421 GACTTAAACTTAGAATGAAGGTTTAAACTGAGAAATGAAGGTTTAAACTGAGAAAT 480

QY 331 gaagggttaaaaactgagaatgaagggttaaaaactgagaatgaagggttcttcttagg 390

Db 481 GAAGGGTTAAAAACTTGAGATGAAGGGTTTAAACTGAGGTGACAGAGTGTTCCTAGG 540

QY 391 ggagcgtgaaaaaccgagac 450

Db 541 GGGACGGTAGAAAAACCCCGCAAGCAGACAGACTCTTAAACAAATATGACGTAGAGAAC 600

QY 451 tctatgttacaagcctaataatgacaagtaccatggagagtgagtgagtgagtgagtgag 510

Db 601 TCTATGTTTACAAAGCTAATAATGACAAGTTACCATGGAGAGTGCGTGTATATAGGCATACG 660

QY 511 ccagaagaataagctaaaaaaattattgaagctcttgacgcaaaagacacacacacacacacacac 570

Db 661 CCAGAAGATAGCTTAAAAAATAATTTATGACGNTCTTCGCCCAAAAGACATGAATTACAA 720

QY 571 caacgaatgagaagttatct 591

Db 721 CAACGAATGAGAACTTATCT 741

RESULT 2

STREMM24

LOCUS

DEFINITION

2131 bp

DNA

S.pyogenes type 24 M-protein gene, BCT complete cds.

BASE COUNT		1153 a	463 c	532 g	485 t
ORIGIN					
/label=Prot-enn5 /product="Enn5_8193 protein" /protein_id="AAA50855.1" /db_xref="GI:437192" /translation="MTROTNNKNSLRKLTGTASVAVALTVLGAFNQTVEVKAEGY RASGTINSENISKLYDENSKLVEERAEFLIDKIDKLEELAKEKKAEDKAEDVERHY LRKLOEHEKEEROKNLEELEROGKREIDKRYOEQKQOQOQLETEKQKLTAEANQVSE ASKGLSNDLEASRAAKKELEAKHOKLEADHOVSETSRKGLSRDLEASREANKKVTSE LTKQAQUSALEESKKLTEKEKAELOAKLEAEAKFEQALKAQAEELAKLRAGKVSDS QTPDAKPGNKVYPGKGQAPQAGAPNQNKMPKKEKQLPSTGETANPEPTAAALATVM ATAGVAAYVVRKEEN"					
Query Match		14.8%	Score	171.8	DB 3; Length 2633;
Best Local Similarity		91.5%	Pred. No.	3.4e-22;	
Matches 182; Conservative		0;	Mismatches	17;	Indels 0; Gaps 0;
QY	171	tgctaaagagaactacgtgattcgccgctgactaggggtacaataataatgacccgcaaaag	230		
DB	30	TGTCAACTACTAATGAAGTTAGTCAACCGCTGACTAGGGGTACAATAAATGACCCGCAAAAG	89		
QY	231	agcaaaagaagctcttgacaagttagcagtagaagaaacacatgacttaaaaactaagaatga	290		
DB	90	AGCAAAAGAAGACTCTTGACAAGTAGTACGTAGAAAACCATGACTTAAAAACTAAGAATGA	149		
QY	291	aggggtaaaaaactgagaatgaaggggttaaaaaactgagaatgaaggggttaaaaaactgagaa	350		
DB	150	AGGGTTAAAAACTGAGAATGAAGGGTTAAAAACTGAGAATGAAGGGTTAAAAACTGAGAA	209		
QY	351	tgaaggggttaaaaaactgag	369		
DB	210	TGAAGGGTTAAAAAACTGAG	228		
RESULT 5 AF288752 LOCUS DEFINITION Streptococcus pyogenes isolate 3148-00 M BCT 02-AUG-2000 partial cds. AF288752 AF288752.1 GI:9652381 ACCESSION VERSION KEYWORDS SOURCE Streptococcus pyogenes. Streptococcus pyogenes Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae Streptococcus. 1 (bases 1 to 817) Beall,B.W. 5' emm sequence emm3.2 Unpublished 2 (bases 1 to 817) Beall,B.W. Direct Submission Submitted (21-JUL-2000) Respiratory Diseases Branch, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA 30333, USA					
FEATURES source location/Qualifiers 1..817 /organism="Streptococcus pyogenes" /isolate="3148-00" /db_xref="taxon:1314" /note="blood isolate recovered from patient in New York population based surveillance; T non typeable; opacity factor negative group: A" <1..>817 /gene="emm" <1..>817 /gene="emm" /note="emm3.2" /codon_start=1					
gene CDS					


```

1..1620
/gene="emm3.1"
/codon_start=1
/transl_table=1
/product="M3 protein"
/protein_id="AA06960.1"
/db_xref="GI:1263023"
/translation="MAKNTNRHVSRKSKTGTASVAVALTVLGTVGAGTVKADAR
SVNGEPFRVHKLVKNIENLLDQVLYTKHNSVOYNQNRGLDLROKAEYLRKGLD
WAERLLQKNGEDVGVKLVGFAFKYSEKVEKKEKIDKKEKEYODLDKDFPLAK
GYVLSDRQHQEELSEKKEKVFATAYGVQISELFTVKQKVESTMQDITPKQNRVSLQ
EQEYLAITKNAREKDFEALAAADKQKLEAKIDETKLKAEKDFEALNNLRKQIAH
NEYAKLAKERDQIQKLEEQRIIDASRKGTRADPEAVQAQKATEALNNLRKQIAE
VTEQGIIDASRKGTRADLEAVQAQKAEVALQLQEQNRTSEASRKGRLRRDLSAR
EAKQVEKALIGANSKLAALBEKLNKELESEKLTKEKEAEQAKLEAEAKLEQALQ
QAELKALRAGKASDSQIPDTKPGNKAVPGKGQAPGATKPNQNKAPMKETTKQLPST
GETANPFPTAALTYMATGYAAVAVKVKKEEN"

```

Query Match	Score	DB 3:	Length
Best Local Similarity	63.4%	Pred. No. 1.2e-19;	
Matches	242;	Conservative	0; Mismatches 140; Indels 0; Gaps 0;
747 tgsatccttggtagatcaggtgtacacaaattatatactaaacataataatagtaattaccaca	806		
180 TGAGAACTTGTTAGATCAGGTTACACAATTATATATCTAAACATATATAGTAATTACCAACA	239		
807 atataatgcacaagctggcagacttgacctgagacaaagcgtgaatctctaaagggcct	866		
240 ATATAATGCACAAGCTGGCAGACTTGACCTGAGACACAAAGGCTCAATATCTAAAGGCT	299		
867 taatgattggctcagagcgtgttacaagagtttaataatcatgattcgacataggtctca	926		
300 TAAATGATGGCTGAGAGGCTTGTTTACAAGAGTTTAATCGAAGATGTAATAAAGTTT	359		
927 gacagatactctgysaaaagtacaagaacgtgctgacagtttgagatagdaaaacaatc	986		
360 AGGTAAGTGGCTTTTGAAGAAAGATGATTATAGAAAGGAGGTTTAAAGACCTTAAGGAAA	419		
987 gttaaaacttaagaatagtgacttaagttttaaataataaagcgttaaaagatcataatga	1046		
420 AATATAGACAAAAGGAAAAGGAATATCAGGACCTTAGATTAAGATTTTTCACCTTGCCCAACA	479		

	047	CyattTaaCtGaaagayttTgagCaacgcTcaadgaagaaaatcaccgataaaaaatgtatataaacctact	1108
Ddb	480	GGGGTAGTTATTCAGATAAAGACATCAACAAGAAGAACTAGAGAAGAAAAGAAAAAGT	539
Oy	1107	atctgaaaaagctagttaaatt	1128
Ddb	540	TACAGAAGCAACTGCTAAAGTT	561
RESULT	7		
STREMM3	LOCUS	STREMM3 1465 bp DNA	. BCT
	DEFINITION	S. pyogenes DNA for M 3 protein, partial sequence.	01-FEB-2000
	ACCESSION	D14415	
	VERSION	D14415.1 GI:511939	
	KEYWORDS	M 3 protein.	
	SOURCE	Streptococcus pyogenes (strain:C203 (ATCC 12384))	DNA.
	ORGANISM	Streptococcus pyogenes Streptococcus pyogenes Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.	
REFERENCE	AUTHORS	1 {bases 1 to 1465) Katsukawa,C.	
	TITLE	Cloning and nucleotide sequence of the type 3 M protein gene (emm3) consisting of an N-terminal variable portion and C-terminal conserved C-repeat regions: Relation to other genes of Streptococcus pyogenes	
JOURNAL	PREFERENCE	J Jpn. Assoc. Infect. Dis. 68, 698-705 (1994)	
		2 {bases 1 to 1465}	


```
/db_xref="SPTREMBL:Q54835"
/translation="MAKNNNRHSYSLRKLKGTASVAVALTVIGTGLVAGQTVKADAR
SVNGEPRHRYKNEIENLDQVOTLYNKHNSYQOYSAQGRKLDLRQRAEYLGKND
WAERLLQELAGVKKVAFKDDIKKEVKEKIDKKEKIDKKEKIDKKEKIDKKEKIDK
GYVLSDRHQHQLKEKKEKVVATATKVGQI SEELTVKQVSTMODLTERKQVRSQ
EQELATTQNAKEDFELAAANAADKQLEAKIADLETKLKEAKEDFELAAALGHQHAH
NEYQAKLEAKDDQIKOLEKQKIDLSAKRKTDLEAVRQAKKATEAELENNLKADLAK
VTEQKQILDASRKGTARDLEAVRQAKOAEALAKOLEEONRSEASRKLGRDLDSR
BAKQVQKDLNLTAEKDKVKEEKI SDASRQGLRRDLDAAREAKKQVEKALEEANSK
LAALKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK
OIPDTKGNKAVPKGQAPQAGTKPNQKAPKETKQLPSTGETANPPTAAALTVM
ATAGVAAVVRKEEN"
gene
1..1746
/mat_peptide
/gene="emm3"
124..1743
/gene="emm3"
/product="M3 protein"
BASE COUNT 742 a 292 c 368 g 344 t
ORIGIN
13.4%; Score 154.8; DB 3; Length 1746;
Best Local Similarity 62.8%; Pred. No. 4.4e-19;
Matches 240; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 747 tggtaacttggtagatcgaggtgacacattatatacaataatagtaattaccaca 806
DB 180 TGAGAACTTGTAGTAGAGTACAGTACACAAATATATATAAACATATAATACCA 239
QY 807 ataatgcacagctggcagacttgacctgagacaaagggctgatatctaaaagcct 866
DB 240 ATATAGTGCACAGCTGGCAGACTTGACCTCAGACAAAGAGCTGAATATCTAAAAGGCCT 299
QY 867 taatgattggctgagagctgttacaagagtttaaatatcagatgcgcactaggtctca 926
DB 300 TAATGATTGGCTGAGAGCTGTTACAGAGTTAAATGCGAGAGATGTAAAAAAGTTT 359
QY 927 gacagatctctggaaaaagtagaacgtgctgcagaagtttgatagatagaacaatac 986
DB 360 AGGTAAAGTGGCTTTGAAAGAGATGATTTAGAAAGGAGGTAAAGAACTTAAGGAAA 419
QY 987 gttaaaacttaagaatagtagtacttaagttttaataataaagcgtttaaagaatcata 1046
DB 420 AATAGACAAAAGGAAAGGATATACAGGACTTAGATAAAGATTTTGACTTGGCCAAACA 479
QY 1047 tgagttaactgaagagttgagtaactgaagagaaactacgttaaaatgataaatacact 1106
DB 480 GGGGTATGTTTATCATGATAAAGACATCAACAGAACTAGAGAAGAAAGAAAGAAAGT 539
QY 1107 atctgaaaaagctagtaaaatt 1128
DB 540 TACAGAGCACTGCTAAGTT 561
RESULT 9
S79254
LOCUS S79254 229 bp DNA BCT 30-NOV-1995
DEFINITION emm-M protein [Streptococcus pyogenes, M1a, Genomic, 229 nt].
ACCESSION S79254
VERSION S79254.1 GI:1087104
KEYWORDS
SOURCE Streptococcus pyogenes M1a.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 229)
AUTHORS Penney,T.J., Martin,D.R., Williams,L.C., de Malmarche,S.A. and Bergquist,P.L.
TITLE A single emm gene-specific oligonucleotide probe does not recognise all members of the Streptococcus pyogenes M type 1
JOURNAL FEMS Microbiol. Lett. 130 (2-3), 145-149 (1995)
MEDLINE 95377598
REMARK GenBank staff at the National Library of Medicine created this
```

```
entry [NCBI gibbsq 169934] from the original journal article.
This sequence comes from Fig. 1.
Location/Qualifiers
1..229
/organism="Streptococcus pyogenes"
/db_xref="taxon:1314"
2..229
/partial
/genes="emm"
/notes="M protein"
2..229
/partial
/genes="emm"
/notes="This sequence comes from Fig. 2"
/codon_start=1
/product="M protein"
/protein_id="AAB35162.1"
/db_xref="GI:1087105"
/translation="SVAVRLTVLGA GFANOTEVKANGDGNPREVIEDLAANPAIQNI
RLRHNKDLKARLENAMVAGRDFKRAEELEK"
BASE COUNT 84 a 32 c 59 g 54 t
ORIGIN
13.0%; Score 151; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacggtgtagtgaatcctcctaggaagttatagaagatcttgcagcaacaatcccccaata 657
DB 65 AACGGTGTAGTGTAACTCTAGGGAAGTTATAGAAGATCTTGCAGCAACAATCCCGCAATA 124
QY 658 caaatatatactttacgtccagaaaaaaggacttaaaagcgaatagagaatgcaatg 717
DB 125 CAAATATACGTTTACGTCACGAAACAAAGGACTTAAAGCGAGATTAGAGTAATGCAATG 184
QY 718 gaagttgaggagaagatgttttaagagagctg 748
DB 185 GAAATTGCGAGAGAGATTTTAAGAGAGCTG 215
RESULT 10
S79254
LOCUS S79254 237 bp DNA BCT 16-MAY-1996
DEFINITION Streptococcus pyogenes M type 1 (emm1) gene, partial cds.
ACCESSION U11940
VERSION U11940.1 GI:533557
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 237)
AUTHORS Whatmore,A.M.
TITLE Sequence Analysis of the Emm-Like Gene Family of Streptococcus Pyogenes
JOURNAL Thesis (1993) Microbiology, University of Newcastle Upon Tyne
REFERENCE 2 (bases 1 to 237)
AUTHORS Whatmore,A.M., Kapur,V., Sullivan,D.J., Musser,J.M. and Kehoe,M.A.
TITLE Non-congruent relationships between variation in emm gene sequences and the population genetic structure of group A streptococci
JOURNAL Mol. Microbiol. 14 (4), 619-631 (1994)
MEDLINE 95198537
REFERENCE 3 (bases 1 to 237)
AUTHORS Whatmore,A.M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1994) EMBL Data Library By: A. M. Whatmore, Microbiology, University of Newcastle Upon Tyne, Framlington Place, Newcastle Upon Tyne, UK, NE2 4HH
Location/Qualifiers
1..237
/organism="Streptococcus pyogenes"
/strain="M type 1"
```

gene	/isolate="NCTC8198"		/db_xref="taxon:1314"	
CDS	1..237		/gene="emml"	
	<1..>237		/gene="emml"	
	/codon_start=1		/transl_table=11	
	/protein_id="AAA9556.1"		/db_xref="GI:53358"	
	/translation="KTGTASVAVALTVLGAGFANOTEVKANGDNPREVIEDLAANNP AIQNTLRHENKDKLARLENAMVAGRDFKRAEL"		54 t	
BASE COUNT	86 a	35 c	62 g	54 t
ORIGIN				
Query Match	13.0%; Score 151; DB 3; Length 237;			
Best Local Similarity	100.0%; Pred. No. 2.4e-18;			
Matches 151; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	598	aacggtgatgtaactcctaggaagtattatagaagattcttcgacaaacaatcccgcaata	657	
DBD	79	RACGGTGATGCTAATCTCTAGGGAAGTTTATAGAAGATCTTCACGACCAACAATCCCGCAATA	138	
QY	658	caaaatatactttacgtccagcaaaaacgaagacttaaacgagattagagaatgcaatg	717	
DBD	139	CAAAATATACGTTTACGTCACGAAAACGAAGACCTAAAGCGAGATTAGAGAATGCAATG	198	
QY	718	gaagttgcagggaagagattttaagagagctg	748	
DBD	199	GAAGTTGCAGGAAGAGATTTTAAGAGAGCTG	229	
RESULT 11				
SP020094	300 bp		DNA	
LOCUS	SP020094		BCT	
DEFINITION	Streptococcus pyogenes M1.0 protein gene, partial cds.		05-JAN-1996	
ACCESSION	U20094			
VERSION	U20094.1		GI:643545	
KEYWORDS	Streptococcus pyogenes.			
SOURCE	Streptococcus pyogenes.			
REFERENCE	Musser, J.M., Kapur, V., Szeto, J., Pan, X., Swanson, D.S. and Martin, D.R.		Streptococcaceae; Streptococcus pyogenes	
AUTHORS	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.			
TITLE	1 (bases 1 to 300)			
JOURNAL	Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease			
MEDLINE	Infect. Immun. 63 (3), 994-1003 (1995)			
REFERENCE	9512752			
AUTHORS	2 (bases 1 to 300)			
TITLE	Direct Submission			
JOURNAL	Submitted (19-JAN-1995)			
FEATURES	College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
source	Location/Qualifiers			
CDS	1..300		/organism="Streptococcus pyogenes"	
	/isolate="MGAS 2116"		/db_xref="taxon:1314"	
	<1..>300		/codon_start=1	
	/transl_table=11		/product="M1.0 protein"	
	/protein_id="AAA85107.1"		/db_xref="GI:643546"	
	/translation="TVLGAGFANOTEVKANGDNPREVIEDLAANNPQNTIKLRHEN KOLKARLENAMVAGRDFKRAELEKAKQALEIDQKDLTKLKLQDDYDLAKEST"		54 t	
BASE COUNT	126 a	45 c	67 g	62 t
ORIGIN				

REFERENCE 2 (bases 1 to 1494)
 AUTHORS Harbaugh,M.P., Podbielski,A., Huegel,S. and Cleary,P.P.
 TITLE Nucleotide substitutions lead to antigenic variation in the
 V-terminal non-repeat domains of group A streptococci M1 protein
 JOURNAL Mol. Microbiol. 7 (1993) In press
 REFERENCE 3 (bases 1 to 1494)
 AUTHORS Harbaugh,M.P., Podbielski,A., Hugl,S. and Cleary,P.P.
 TITLE Nucleotide substitutions and small-scale insertion produce size and
 antigenic variation in group A streptococcal M1 protein
 JOURNAL Mol. Microbiol. 8 (5), 981-991 (1993)
 MEDLINE 93360826
 COMMENT Product subunit structure is a dimer. See also x62132 for sequence.
 FEATURES
 source
 1. .1494
 /organism="Streptococcus pyogenes"
 /isolate="prague 11-29/58"
 /db_xref="taxon:1314"
 10. .14
 /evidence=experimental
 31. .1485
 /gene="emml"
 31. .1485
 /gene="emml"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="M protein type 1"
 /protein_id="CAA44062.1"
 /db_xref="GI:311758"
 /db_xref="SPTREMBL:Q10372"
 /translation="MAKNNTRHYSRLKLTGTASVAVALTVLGAFANOTEVKANGD
 GNPREVIEDLAANNPAIQNIRLHNENKDKLKALENAMVAGDFKRAELEKAKOALE
 LEKELEKKKALELAIDQASODYNRANVLEKELDTITREQIENRNLGNKLELDQLS
 SEKQLITERAKLEERKQISDASROSLRDLDAAREAKQKVEKDLANLPAELDKVKED
 KOISDASROGLRDLDAAREAKQKVEKDLANLPAELDKVKEDQISDASROGLRDLDA
 ASRAKQKVEKDLANLPAELDKVKEDQISDASROGLRDLDAAREAKQKVEKDLANL
 LAKQAELEKLRGAKSDSQTPDKPGNKVPGKQAPQAGTFPNKAPMKETKRLQ
 PSTGETANPFETRAALTVATGAAGVAAVVRKEEN"
 BASE COUNT 624 a 268 c 319 g 283 t
 ORIGIN

Query Match 13.0%; Score 151; DB 3; Length 1494;
 Best Local Similarity 100.0%; Pred. No. 2.2e-18;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 598 aacggtgatgtaactcctaggaagttagaagatcttcagcaaacacacccgcaata 657
 DB 154 AACGGTGATGTAATCTTAGGGAAGTTATAGAGATCTTGCGACCAACATCCGCAATA 213
 QY 658 caaataatagctttacgtcacgaaacaaaggacttaaaagcgagattagagaatgcaatg 717
 DB 214 CAAAATATAGCTTTAGCTCAGCAAAACAAAGGACTTAAAGCGGAGATTAGAGAAATGCAATG 273
 QY 718 gaagttgcaggaagagatttaagagagctg 748
 DB 274 GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG 304
 RESULT 14
 SPEM1 1708 bp DNA BCT 14-MAR-1996
 LOCUS S.pyogenes emml gene for type 1 group A M protein.
 DEFINITION X07933 X07860
 ACCESSION X07933.1 GI:47358
 VERSION emml gene; M protein.
 KEYWORDS Streptococcus pyogenes.
 SOURCE Streptococcus pyogenes.
 ORGANISM Streptococcus pyogenes.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 1708)
 AUTHORS Hannes-Fritz,E.

Direct Submission
 Submitted (15-JUN-1988) Hannes-Fritz E., University of Minnesota,
 Department of Microbiology, Box 196, Mayo Building, Minneapolis, MN
 55455
 REFERENCE 2 (bases 1 to 1708)
 AUTHORS Hannes-Fritz,E., Kraus,W., Burdett,V., Dale,J.B., Beachey,E.H. and
 Cleary,P.
 TITLE Comparison of the leader sequences of four group A streptococcal M
 protein genes
 JOURNAL Nucleic Acids Res. 16 (10), 4667-4677 (1988)
 MEDLINE 88247768
 FEATURES
 Location/Qualifiers
 1. .1708
 /organism="Streptococcus pyogenes"
 /strain="CS130"
 /db_xref="taxon:1314"
 /clone="pbf82"
 921. .963
 /note="pot.promoter region"
 1015. .1019
 1028. .>1708
 /codon_start=1
 /transl_table=11
 /product="M protein precursor"
 /protein_id="CAA30757.1"
 /db_xref="GI:47359"
 /db_xref="SPTREMBL:Q10371"
 /translation="MAKNNTRHYSRLKLTGTASVAVALTVLGAFANOTEVKANGD
 GNPREVIEDLAANNPAIQNIRLHNENKDKLKALENAMVAGDFKRAELEKAKOALE
 DOKDLTKLEKQDYDLAKESTWDRQLEKELEKEKLEKELEKELEKELEKELEKELEKELE
 LEKELEKKKALELAIDQASODYNRANVLEKELETTITREQIENRNLGNKLELDQLS
 SEKEQLTIR"
 sig_peptide 1028. .>1708
 mat_peptide 1151. .>1708
 repeat_region 1427. .1525
 repeat_region 1522. .1609
 repeat_region /rpt_type=DIRECT
 repeat_region /rpt_type=INDIRECT
 BASE COUNT 640 a 310 c 331 g 427 t
 ORIGIN
 Query Match 13.0%; Score 151; DB 3; Length 1708;
 Best Local Similarity 100.0%; Pred. No. 2.2e-18;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 598 aacggtgatgtaactcctaggaagttagaagatcttcagcaaacacacccgcaata 657
 DB 1151 AACGGTGATGTAATCTTAGGGAAGTTATAGAGATCTTGCGACCAACATCCGCAATA 1210
 QY 658 caaataatagctttacgtcacgaaacaaaggacttaaaagcgagattagagaatgcaatg 717
 DB 1211 CAAAATATAGCTTTAGCTCAGCAAAACAAAGGACTTAAAGCGGAGATTAGAGAAATGCAATG 1270
 QY 718 gaagttgcaggaagagatttaagagagctg 748
 DB 1271 GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG 1301
 RESULT 15
 SPEM1 251 bp DNA BCT 16-MAY-1996
 LOCUS Streptococcus pyogenes M type 68 (emml) gene, partial cds.
 DEFINITION U11997
 ACCESSION U11997.1 GI:533671
 VERSION U11997.1
 KEYWORDS Streptococcus pyogenes.
 SOURCE Streptococcus pyogenes.
 ORGANISM Streptococcus pyogenes.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 251)
 AUTHORS Whatmore,A.M.

TITLE Sequence Analysis of the Emm-Like Gene Family of Streptococcus
Pyogenes
JOURNAL Thesis (1993) Microbiology, University of Newcastle Upon Tyne
REFERENCE 2 (bases 1 to 251)
AUTHORS Whatmore,A.M., Kapur,V., Sullivan,D.J., Musser,J.M. and Kehoe,M.A.
TITLE Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci
JOURNAL Mol. Microbiol. 14 (4), 619-631 (1994)
MEDLINE 95198537
REFERENCE 3 (bases 1 to 251)
AUTHORS Whatmore,A. M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1994) Embl Data Library By: A. M. Whatmore,
Microbiology, University of Newcastle Upon Tyne, Framlington Place,
Newcastle Upon Tyne, UK, NE2 4HH
FEATURES
source
1..251
Location/Qualifiers
/organism="Streptococcus pyogenes"
/strain="M type 68"
/isolate="R80/2076"
/db_xref="taxon:1314"
1..251
/gene="emm1"
<1..>251
/gene="emm1"
/codon_start=1
/transl_table=11
/protein_id="AA099613.1"
/db_xref="GI:533672"
/translation="LRKLKGTASVAVALTVLGAGFANQTEVKANGDGNPREVIEDLP
ANPAIQIRLRHKNKDLKARLENAMVAGRDFKRAEELS"
BASE COUNT 92 a 38 c 62 g 59 t
ORIGIN

Query Match 12.9%; Score 149.4; DB 3; Length 251;
Best Local Similarity 99.3%; Pred. No. 4.8e-18;
Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 598 aacggatgagtgatctcctaggggaagttatagaagatcttgcagcaacaatccgcgaata 657
|||||
DB 91 AACGGTGTATCTCTAGGGAAGTTATAGAAGATCTTCCAGCAACAATCCCGCAATA 150
|||||
QY 658 caaataatcgtttacgtcacgaaacagagacttaaaagcgagattagagaatgcaatg 717
|||||
DB 151 CAAATAATACGTTTACGTCACGAAACAAAGGACTTAAAGCGAGATTAGAGATGCAATG 210
|||||
QY 718 gaagttgcagagagagattttaagagagctg 748
|||||
DB 211 GAAGTTGCAGGAGAGATTTTAAAGAGCTG 241
|||||

Search completed: May 13, 2001, 04:55:34
Job time: 10569 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 02:42:00 ; Search time 119.14 seconds
(without alignments)
5674.154 Million cell updates/sec

Title: US-09-151-409-15

Perfect score: 1158

Sequence: 1 gcattgatgtcgcgactag.....aggcagctaagtaaaagctt 1158

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_0401.*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1158	100.0	1158	20	X33103
2	419	36.2	741	21	A57894
3	417.4	36.0	741	15	Q45215
4	328.2	28.3	822	15	Q45216
5	328.2	28.3	822	21	A57895
6	301.8	26.1	741	15	Q45219
7	301.8	26.1	741	21	A57898
8	301.8	26.1	1029	15	Q45221
9	301.8	26.1	1029	21	A57900
10	241	20.8	861	15	Q45214
11	241	20.8	861	21	A57893

12	239.4	20.7	522	15	Q45220
13	239.4	20.7	522	21	A57899
14	124.4	10.7	561	15	Q45217
15	124.4	10.7	561	21	A57896
16	106.4	9.2	1617	18	T49317
17	105.6	9.1	301	6	N50341
18	76.2	6.6	408	15	Q45223
19	76.2	6.6	408	21	A57902
20	68	5.9	2943	21	A70229
21	64.2	5.5	3579	21	A70099
22	63.2	5.5	1998	21	A70212
23	61	5.3	6644	20	X33181
24	61	5.3	7372	20	X33182
25	61	5.3	7797	20	X33180
26	61	5.3	7996	20	X33184
27	57.6	5.0	1686	16	Q87587
28	56.4	4.9	1039	20	X61812
29	56.4	4.9	1125	20	X61811
30	55.8	4.9	35515	20	X20252
31	55.8	4.8	3399	17	T05868
32	54.8	4.7	855	15	Q45161
33	53.6	4.6	7458	21	A70106
34	53.2	4.6	204	15	Q45222
35	53.2	4.6	204	21	A57901
36	53.2	4.6	918	15	Q45218
37	53.2	4.6	918	21	A57897
38	53.2	4.6	2658	21	A70213
39	53	4.6	765	15	Q45160
40	52.4	4.5	3738	21	A70178
41	51.8	4.5	4590	7	M60472
42	50.6	4.4	7503	21	A70206
43	49.2	4.2	5940	21	A70105
44	49	4.2	1664976	19	V21209
45	48.4	4.2	11922	21	A70187

ALIGNMENTS

RESULT 1

X33103
ID X33103 standard; DNA; 1158 BP.

XX AC X33103;

XX XX 23-JUN-1999 (first entry)

XX DE Streptococcal hexavalent M protein vaccine encoding DNA.

XX XX Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A;
KW immunogenic; immune response; pathogen; fusion protein; ss.
XX OS Streptococcus sp.

XX PN WO9913084-A1.

XX PD 18-MAR-1999.

XX PF 14-SEP-1998; 98WO-US19100.

XX PR 12-SEP-1997; 97US-0058635.

XX XX (IDVA-) ID VACCINE.

XX PA Dale JB;

XX PI WPI; 1999-215066/18.

XX DR P-PSDB; Y04368.

XX DR S. pyogenes hybrid

XX DR TetraValent-C repe

XX PT S. pyogenes hybrid

XX PS S. pyogenes hybrid

XX PS Recombinant M24-M5

XX PS S. pyogenes hybrid

XX PS Immunogenic fusion protein derived from group A streptococci

XX PS Example 1; Fig 7; 50pp; English.

Recombinant M24-M5
S. pyogenes hybrid
Recombinant M24-M5
S. pyogenes hybrid
Type-6 M-protein e
Sequence of a port
M24-M5-M6-M19-M3-M
S. pyogenes octava
Plasmodium falcipa
Plasmodium falcipa
Plasmodium falcipa
Base sequence of t
Base sequence of t
Cowpox virus bsr f
Base sequence of t
DNA encoding leuco
B. burgdorferi ant
B. burgdorferi ant
Borrelia burgdorfe
Chicken leucocytoz
Sequence of fragme
Plasmodium falcipa
M24-M5-M6-M19 hybr
S. pyogenes hybrid
Recombinant M24-M5
S. pyogenes hybrid
Plasmodium falcipa
Sequence of a frag
Plasmodium falcipa
Sequence encoding
Plasmodium falcipa
Plasmodium falcipa
Methanococcus jann
Plasmodium falcipa

The present invention describes an immunogenic fusion polypeptide (I) that stimulates an immune response against group A streptococci (GAS). (I) comprises: (a) at least 2 peptides (Ia) from GAS, at least 10 amino acids long and able to induce a specific immune response; and (b) a peptide (Ib), C-terminal to (Ia) that protects the immunogenicity of the component described in (a) but is not essential for stimulation of the immune response. Vaccines containing (I) are used to protect against GAS infections, specifically those caused by Streptococcus pyogenes, e.g. pharyngitis, pyoderma, toxic shock syndrome, deep tissue invasion, sepsis and acute rheumatic fever. (I) have improved immunogenicity and do not generate antibodies that are cross-reactive with human tissues. The present sequence encodes a hexavalent M protein vaccine.

Sequence 1158 BP; 474 A; 165 C; 255 G; 264 T; 0 other;

Query Match 100.0%; Score 1158; DB 20; Length 1158;
Best Local Similarity 100.0%; Pred. No. 4,1e-244;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcatgttcgactaggtctcagacagatactctggaaaagtacaaagacgtgct 60
Db 1 gcatgcatgttcgactaggtctcagacagatactctggaaaagtacaaagacgtgct 60
QY 61 gacaagttagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttta 120
Db 61 gacaagttagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttta 120
QY 121 aataaagcgttaaaagatcataatgatgagtttaactgaaggttgagtaactgaagag 180
Db 121 aataaagcgttaaaagatcataatgatgagtttaactgaaggttgagtaactgaagag 180
QY 181 aaactacgttgatccgcgtgactaggtgacataataataatgacccgcaagagcaaa 240
Db 181 aaactacgttgatccgcgtgactaggtgacataataataatgacccgcaagagcaaa 240
QY 241 gctcttcaagaagtatgactagaaacacatgacttaaaactaaagaatgaagggttaa 300
Db 241 gctcttcaagaagtatgactagaaacacatgacttaaaactaaagaatgaagggttaa 300
QY 301 actgagaatgaagggtttaaactgagaatgaagggtttaaactgagaatgaagggtta 360
Db 301 actgagaatgaagggtttaaactgagaatgaagggtttaaactgagaatgaagggtta 360
QY 361 aaactgaggtcagacagaggtttcctaggggacgtgagaaacccgcagacaaacaga 420
Db 361 aaactgaggtcagacagaggtttcctaggggacgtgagaaacccgcagacaaacaga 420
QY 421 gaactcttaacaagtatgacgtagagaactctatgttacaagctaaatgacaagtta 480
Db 421 gaactcttaacaagtatgacgtagagaactctatgttacaagctaaatgacaagtta 480
QY 481 ccattgagagtgctttactactagcagacacccagagataagctaaataattattgac 540
Db 481 ccattgagagtgctttactactagcagacacccagagataagctaaataattattgac 540
QY 541 gatcttgacgcaaaagaacatgaattacaacacagaaatgagaagtattctctgcggaac 600
Db 541 gatcttgacgcaaaagaacatgaattacaacacagaaatgagaagtattctctgcggaac 600
QY 601 ggtatggttaactcctagggagttatagaagattcttcagcaacaacatcccgcaataca 660
Db 601 ggtatggttaactcctagggagttatagaagattcttcagcaacaacatcccgcaataca 660
QY 661 aatatcgtttacgtcacgaaacaagacttaaaagcgagattagagaatgcaatggaa 720
Db 661 aatatcgtttacgtcacgaaacaagacttaaaagcgagattagagaatgcaatggaa 720
QY 721 gtgtcaggaagagattttaagagagctgtggtaccctgttagatcaggtttacaattatat 780
Db 721 gtgtcaggaagagattttaagagagctgtggtaccctgttagatcaggtttacaattatat 780
QY 781 actaaacataatagtaattaccacaacataataatgacaaagctggcagacttgacctgaga 840

Db 781 actaaacataatagtaattaccacaacataataatgacaaagctggcagacttgacctgaga 840
QY 841 caaagcctgaatatactaaagccttaataatgattggctgagagcgtgttacaagagttta 900
Db 841 caaagcctgaatatactaaagccttaataatgattggctgagagcgtgttacaagagttta 900
QY 901 aatcagatgcgcgactaggtctcagacagatactctgaaaaagtacaaagacgtgct 960
Db 901 aatcagatgcgcgactaggtctcagacagatactctgaaaaagtacaaagacgtgct 960
QY 961 gacaagtttgagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttta 1020
Db 961 gacaagtttgagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttta 1020
QY 1021 aataaagcgttaaaagatcataatgatgagtttaactgaaggttgagtaactgaagag 1080
Db 1021 aataaagcgttaaaagatcataatgatgagtttaactgaaggttgagtaactgaagag 1080
QY 1081 aaactacgttaaaatgataatcactatctgaaaaagctagtaaaattcaagaattagag 1140
Db 1081 aaactacgttaaaatgataatcactatctgaaaaagctagtaaaattcaagaattagag 1140
QY 1141 gacagtaagtaaaagcctt 1158
Db 1141 gacagtaagtaaaagcctt 1158

RESULT 2
A57894
ID A57894 standard; DNA; 741 BP.
XX A57894;
XX
XX
DT 10-OCT-2000 (first entry)
XX
DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:3.
XX
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX
OS Streptococcus pyogenes.
OS Synthetic.
XX
XX US6063386-A.
XX
XX 16-MAY-2000.
XX
XX 15-SEP-1997; 97US-0937271.
XX
XX 16-SEP-1992; 92US-0945954.
XX
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX Lederer JW, Dale JB;
XX
XX WPI; 2000-364475/31.
DR P-PSDB; B031114.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever -
XX
XX Disclosure; Fig 4A-B; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein

CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences A57893-A57902
 CC represent DNAs encoding multivalent hybrid streptococcus pyogenes M
 CC proteins generated in the disclosure of the invention.
 XX
 SQ Sequence 741 BP; 311 A; 106 C; 164 G; 160 T; 0 other;

Query Match 36.2%; Score 419; DB 21; Length 741;
 Best Local Similarity 78.8%; Pred. No. 7.6e-83;
 Matches 585; Conservative 0; Mismatches 0; Indels 156; Gaps 1;

Qy 7 atggtcgactagctctcagacagatactctggaaaaagatacaagaacgtgctgacaag 66
 Db 1 atggtcgactagctctcagacagatactctggaaaaagatacaagaacgtgctgacaag 60
 Qy 67 ttgagatagaacaatacgtttaaacttaagaactagtgacttaagttttaataataa 126
 Db 61 ttgagatagaacaatacgtttaaacttaagaactagtgacttaagttttaataataa 120
 Qy 127 gcttaaaagatcataatgattgaactgaagaggttgagtaagtctaaagagaacta 186
 Db 121 gcttaaaagatcataatgattgaactgaagaggttgagtaagtctaaagagaacta 180
 Qy 187 cgt----- 189
 Db 181 cgtaaaaatgataaatacactatctgaaaaagctagtaaaatacaagaattagaggcagt 240
 Qy 190 ----- 189
 Db 241 aaggctgatcttgaaaaagcattagaagcgccaatgaatttttcaacacgqgattcagct 300
 Qy 190 -----gactccgcgtgactagggt 210
 Db 301 aaaaataaacttagaagcagagaagcgtgacttgaaaggatccgcgtgactagggggt 360
 Qy 211 acaataaatgacccgcaagaagcaaaagactcttgacaagatgagctagaaaaaccat 270
 Db 361 acaataaatgacccgcaagaagcaaaagactcttgacaagatgagctagaaaaaccat 420
 Qy 271 gacttaaaactaagaatgaagggttaaaaactgagaatgaagggttaaaaactgagaat 330
 Db 421 gacttaaaactaagaatgaagggttaaaaactgagaatgaagggttaaaaactgagaat 480
 Qy 331 gaagggttaaaaactgagaatgaagggttaaaaactgagctgcagagaggtttctcagg 390
 Db 481 gaagggttaaaaactgagaatgaagggttaaaaactgagctgcagagaggtttctcagg 540
 Qy 391 ggaacggtgaaaccccgcaagaagcagcagaactctttaaacaagatgacgtagagaac 450
 Db 541 ggaacggtgaaaccccgcaagaagcagcagaactctttaaacaagatgacgtagagaac 600
 Qy 451 tctatgtacaagctaaataatgacaagttaccatggagagtgctgttatactaggcatac 510
 Db 601 tctatgtacaagctaaataatgacaagttaccatggagagtgctgttatactaggcatac 660
 Qy 511 ccagaagaagaataaaaaaattatgacgtatcttgacgcaaaagaacatgaattacaa 570

Db 661 ccagaagaagaataaaaaaattatgacgtatcttgacgcaaaagaacatgaattacaa 720
 Qy 571 caacagaatgagaagttatct 591
 Db 721 caacagaatgagaagttatct 741
 RESULT 3
 Q45215
 ID Q45215 standard; DNA; 741 BP.
 XX
 AC Q45215;
 XX
 DT 02-NOV-1994 (first entry)
 XX
 DE Recombinant M24-M5-M6-M19 gene.
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX
 OS Synthetic.
 XX
 FN W09406421-A.
 XX
 PD 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 DR WPI; 1994-118122/14.
 DR P-PSDB; R50993.
 XX

New immunogenic hybrid proteins derives from streptococcal M proteins
 - induces opsonic antibodies, for protective immunisation against
 against multiple group A streptococci serotypes

Disclosure; Fig 4; 67pp; English.

CC The sequences given in Q45214-23 encode hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC pK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.

Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;

Query Match 36.0%; Score 417.4; DB 15; Length 741;
 Best Local Similarity 78.8%; Pred. No. 1.7e-82;
 Matches 584; Conservative 0; Mismatches 1; Indels 156; Gaps 1;

Tue May 15 07:27:08 2001

Qy	7	atggtcgcgactaggctctcagacagatactctggaagaaagtacaagaacgctgcgcacaag	66
Db	1	atggtcgcgactaggctctcagacagatactctggaagaaagtacaagaacgctgcgcacaag	60
Qy	67	ttgagatagaaacataagcttaaaacttaagaactagcttaagtttttaataataa	126
Db	61	ttgagatagaaacataagcttaaaacttaagaactagcttaagtttttaataataa	120
Qy	127	cgcttaaaagatcataatgatgaggttaactgaagagaggttagtaagtaaaagagaacta	186
Db	121	cgcttaaaagatcataatgatgaggttaactgaagagaggttagtaagtaaaagagaacta	180
Qy	187	cggt-----	189
Db	181	cgtaaaatgatataactactatctgaaaaagcctagtataaaatcaagaattagaggcacgt	240
Qy	190	-----	189
Db	241	aaggtgatcttgaaaaagcattagaaagcgcaatgaattttcaacagcggattcagct	300
Qy	190	-----ggatccgcgtgactagggt	210
Db	301	aaaatcaaaacaccttagaagcagagagaagctgatcttgaaggatccgcgtgactagggt	360
Qy	211	acaataaatgacccgcaaaagagcaaaagaagctcttgcacaagtatgaggtagaaaacat	270
Db	361	acaataaatgacccgcaaaagagcaaaagaagctcttgcacaagtatgaggtagaaaacat	420
Qy	271	gacttaaaactaagaatgaaggtttaaactgagaatgaaggtttaaactgagaat	330
Db	421	gacttaaaactaagaatgaaggttcaaactgagaatgaaggtttaaactgagaat	480
Qy	331	gaaggtttaaactgagaatgaaggtttaaactgagaatgaaggtttaaactgagaat	390
Db	481	gaaggtttaaactgagaatgaaggtttaaactgagaatgaaggtttaaactgagaat	540
Qy	391	gggacggttagaaacccggaagcagcagcaactcttaacaagtatgacgtagagaaac	450
Db	541	gggacggttagaaacccggaagcagcagcaactcttaacaagtatgacgtagagaaac	600
Qy	451	tctatgttacaagctataataagcaagttaccatgagagtgcttatactaggcatacg	510
Db	601	tctatgttacaagctataataagcaagttaccatgagagtgcttatactaggcatacg	660
Qy	511	ccagaagataagctaaaaaattatitgacgactcttgacgcacaaagaaacatgaattacaa	570
Db	661	ccagaagataagctaaaaaattatitgacgactcttgacgcacaaagaaacatgaattacaa	720
Qy	571	caacagaatgagaagttatct	591
Db	721	caacagaatgagaagttatct	741
RESULT 4			
ID	Q45216 standard; DNA; 822 BP.		
XX	Q45216;		
AC	(first entry)		
XX	02-NOV-1994		
DT	Recombinant M24-M5-M6-M19 (linker variant) gene.		
DE	Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;		
XX	M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;		
KW	tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;		
KW	restriction enzyme site; multivalent M protein; immunisation; group A;		
KW	streptococci; rheumatic fever; rheumatic heart disease; humoral;		
KW	antibody; heart tissue; antigen; serotype; mucosal; ss.		
OS	Synthetic.		
XX			

PN	WO9406421-A.
XX	31-MAR-1994.
XX	15-SEP-1993; 93WO-US08703.
XX	16-SEP-1992; 92US-0945954.
PR	(UYTE-) UNIV TENNESSEE RES CORP.
PA	Dale JB, Lederer JW;
PI	WPI; 1994-118122/14.
XX	P-PSDB; R50994.
XX	New immunogenic hybrid proteins derives from streptococcal M proteins
PT	- induces opsonic antibodies, for protective immunisation against
PT	against multiple group A streptococci serotypes
XX	Disclosure; Fig 6; 67pp; English.
PS	The sequences given in Q45214-23 encode hybrid M proteins which
XX	contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
CC	proteins were constructed using fragments of the 5' regions of emm
CC	genes that were amplified by PCR, ligated in tandem and expressed in
CC	pkK223.3. The amplified regions pref. encode protective and not
CC	tissue-cross-reactive epitopes, which can then be linked into one
CC	protein molecule. The recombinant hybrid protein may contain 113
CC	N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC	35 from M19. Each section is linked by 2 amino acids specified by
CC	the respective restriction enzyme sites that were synthesised into
CC	the primers used to specify the PCR product. Multivalent M proteins
CC	such as this may be used for protective immunisation against group A
CC	streptococci, which esp. cause rheumatic fever and rheumatic heart
CC	disease. Humoral antibodies raised against these proteins do not
CC	react with heart tissue antigens but are effective against many
CC	different serotypes. The multivalent proteins may also include
CC	sequences which induce mucosal antibodies and do not require coupling
CC	to an immunogenic carrier.
XX	Sequence 822 BP; 332 A; 136 C; 182 G; 172 T; 0 other;
SQ	

Query Match	28.3%;	Score	328.2;	DB	15;	Length	822;
Best Local Similarity	87.6%;	Pred. No.	5.1e-63;				
Matches	404;	Conservative	0;	Mismatches	3;	Indels	54;
Gaps							
Qy	185	tacgtgacccgcgtgactagggtacaataaataatgacccgcgaagagcaaaagaagctc	244				
Db	362	tccaggatccgcgtgactagggtacaataaataatgacccgcgaagagcaaaagaagctc	421				
Qy	245	ttgacaagtatgactagaaaaccatgacttaaaactaagaatgaaggggttaaaaaactg	304				
Db	422	ttgacaagtatgactagaaaaccatgacttaaaactaagaatgaaggggttaaaaaactg	481				
Qy	305	agaataaaggggttaaaaaactgagaatgaaggggttaaaaaactgagaatgaaggggttaaaaa	364				
Db	482	agaataaaggggttaaaaaactgagaatgaaggggttaaaaaactgagaatgaaggggttaaaaa	541				
Qy	365	ctgag-----gtcgacagagctgttctcctagggggacgg	397				
Db	542	ctgaggtcgcacccaggaaacccagctgttccagtcgacagagtggttctcctagggggacgg	601				
Qy	398	tagaaaaccggaacaaagcagcagaaacttcttaacaagtatgacgtagagaactatgt	457				
Db	602	tagaaaaccggaacaaagcagcagaaacttcttaacaagtatgacgtagagaactatgt	661				
Qy	458	tacaagctaaataatgacaagttaacc-----atggagag	490				
Db	662	tacaagctaaataatgacaagttaacc-----atggagag	721				
Qy	491	tgcgttatactaggcatagcgcagaagaatgaagcttaaaaaataattatgcgactttgacg	550				

Query Match	28.3%	Score 328.2;	DB 21;	Length 822;
Best Local Similarity	87.6%;	Pred. No. 5.1e-63;		
Matches 404;	Conservative	0;	Mismatches 3;	Indels 54;
				Gaps 2;

The invention relates to multivalent immunogenic hybrid group A streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19. The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a multivalent hybrid M protein; and a method for administering a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines and host tissue contain similar antigenic motifs against various serotypes are capable of raising sero-specific not cross-reactive with human heart tissue. Sequences A57893-A57902 represent DNAs encoding multivalent hybrid Streptococcus pyogenes M proteins generated in the disclosure of the invention.

OR P-PSDB; R50997.

CC proteins generated in the disclosure of the invention.
 XX Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;
 SQ

Query Match 26.1%; Score 301.8; DB 21; Length 741;
 Best Local Similarity 68.3%; Pred. No. 2.9e-57;
 Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtcgttatactagcaccagagataagctaaagcagagattgagaaattattgacgatct 545
 |||||
 Db 3 gagagtcgttatactagcaccagagataagctaaagcagagattgagaaattattgacgatct 62
 |||||

QY 546 tgacgcaaaagacatgaattacaacacagaaatgagaattctctgcagacggtga 605
 |||||
 Db 63 tgacgcaaaagacatgaattacaacacagaaatgagaattctctgcagacggtgtt 122
 |||||

QY 606 tggtaatccttagggaggttagaagattctgcagcaaaacatcccgcaatacaaaatat 665
 |||||
 Db 123 tcttagggggacggtagaaaacccgg-----acaaagc 155

QY 666 acgtttacgtcacgaaacaaaggacttaaaagcagagattgagaaatgcaatggaagtgc 725
 |||||
 Db 156 acgagaactcttacaagctagcgtacacataaattgacccgcaaaagcgaagagctct 275

QY 726 aggaagagatttacaacaaataatgcacaaagctggcagacttgacattacacaattatataa 785
 |||||
 Db 216 cttagtcgacgcgtgactaggggtacataaattgacccgcaaaagcgaagagctct 275

QY 786 acataatagtaattacaacaaataatgcacaaagctggcagacttgacattacacaattatataa 845
 |||||
 Db 276 tgacaaagttagctagaaacacatgacttaaaactaagaatgaagggttaaaactga 335

QY 846 ggctgaatatataaagcccttaattggtgctgagagcgtgttacaaggtttaaataat 905
 |||||
 Db 336 gaatgaagggttaaaactgagaatgaagggtttaaactgagaatgaagggtttaaact 395

QY 906 cga-----tgctgcagactaggtctcagacagatactctggaagaagtacagaacgtgc 959
 |||||
 Db 396 tgagccatggtgcagactaggtctcagacagatactctggaagaagtacagaacgtgc 455

QY 960 tgacaagtgtgagatagaacaaactgtaaaacttaagaatagacttaagtttaa 1019
 |||||
 Db 456 tgacaagtgtgagatagaacaaactgtaaaacttaagaatagacttaagtttaa 515

QY 1020 taataaagcgttaaaagataatagtgagtttaactgaagaggttgagtaagtctaaga 1079
 |||||
 Db 516 taataaagcgttaaaagataatagtgagtttaactgaagaggttgagtaagtctaaga 575

QY 1080 gaaactacgtataaataatgataatcactatctgaaaagcgtgtaaaattcaagaattaga 1139
 |||||
 Db 576 gaaactacgtataaataatgataatcactatctgaaaagcgtgtaaaattcaagaattaga 635

QY 1140 ggcacgttaagtaaaagctt 1158
 |||||
 Db 636 ggcacgttaaggtgactctt 654

RESULT 8

Q45221
 ID Q45221 standard; DNP; 1029 BP.
 XX
 AC Q45221;
 XX
 DT 02-NOV-1994 (first entry)
 DE Tetraivalent-C repeat gene.
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX Synthetic.
 OS
 PN WO9406421-A.
 XX
 PD 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 DR WPI; 1994-118122/14.
 XX
 DR P-PSDB; R50999.
 XX
 PT New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX
 PS Disclosure; Fig 11; 67pp; English.
 XX
 CC The sequences given in Q45214-23 encode hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC pK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX
 SQ Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 26.1%; Score 301.8; DB 15; Length 1029;
 Best Local Similarity 68.3%; Pred. No. 3.1e-57;
 Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtcgttatactagcaccagagataagctaaagcagagattgagaaattattgacgatct 545
 |||||
 Db 3 gagagtcgttatactagcaccagagataagctaaagcagagattgagaaattattgacgatct 62
 |||||

QY 546 tgacgcaaaagacatgaattacaacacagaaatgagaattctctgcagacggtga 605
 |||||
 Db 63 tgacgcaaaagacatgaattacaacacagaaatgagaattctctgcagacggtgtt 122
 |||||

QY 606 tggtaatccttagggaggttagaagattctgcagcaaaacatcccgcaatacaaaatat 665
 |||||
 Db 123 tcttagggggacggtagaaaacccgg-----acaaagc 155

QY 666 acgtttacgtcacgaaacaaaggacttaaaagcagagattgagaaatgcaatggaagtgc 725
 |||||
 Db 156 acgagaactcttacaagctagcgtacacataaattgacccgcaaaagcgaagagctct 215

QY 726 aggaagagatttacaagagcgtgactctgttagatcaggtttacacaattatataa 785
 |||||
 Db 216 cttagtcgacgcgtgactaggggtacataaattgacccgcaaaagcgaagagctct 275

QY 786 acataatagtaattacaacaaataatgcacaaagctggcagacttgagacataa 845
 |||||

276 tqacaagtatgagctagaaacacatgacttaataaaactaagaatgaagggttaaaaaactga 335
 846 ggctgaatatctaaagccttaataatgattggctgagagcgtgttacaagagttataatat 905
 336 gaatgaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaaac 395
 906 cga-----tgctgcgactagctcagacagatgactctggaataaagtaacaagaactgc 959
 396 tgagccatgggttcgcgactagctcagacagatgactctggaataaagtaacaagaactgc 455
 960 tgacaagtttgagatagaatacaatacgtttaaaacttaagaatgagcttaacttaagttaa 1019
 456 tgacaagtttgagatagaatacaatacgtttaaaacttaagaatgagcttaacttaagttaa 515
 1020 taataaagcgttaaaagatcataatgagtttaaaacttaagaatgagcttaacttaagttaa 1079
 516 taataaagcgttaaaagatcataatgagtttaaaacttaagaatgagcttaacttaagttaa 575
 1080 gaaactacgtataaaatgataatcactatctgaaaaagctagtaaaattcaagaattaga 1139
 576 gaaactacgtataaaatgataatcactatctgaaaaagctagtaaaattcaagaattaga 635
 1140 ggcacgttaagtaaaagcgt 1158
 636 ggcacgttaagcgtgatctt 654

RESULT 9
 A57900
 ID A57900 standard; DNA; 1029 BP.
 AC
 XX
 XX
 XX
 DT
 DE
 XX
 XX
 KW
 KW
 KW
 KW
 OS
 OS
 OS
 PN
 XX
 XX
 PD
 XX
 XX
 PF
 PF
 XX
 PR
 XX
 XX
 PA
 XX
 PI
 XX
 DR
 DR
 XX
 PT
 PT
 PT
 XX
 PS
 XX
 CC
 CC
 CC
 CC
 CC
 CC

10-OCT-2000 (first entry)
 S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:16.
 Multivalent hybrid M protein; group A streptococcus; serotype;
 immunogenic; sero-specific antibody; streptococcal infection;
 cross reactivity; vaccine; acute rheumatic fever; ARF;
 rheumatic heart disease; streptococcal pharyngitis; strep throat;
 pneumonia; ds.
 Streptococcus pyogenes.
 Synthetic.
 US6063386-A.
 16-MAY-2000.
 15-SEP-1997; 97US-0937271.
 16-SEP-1992; 92US-0945954.
 (UYTE-) UNIV TENNESSEE RES CORP.
 Lederer JW, Dale JB;
 WPI: 2000-364475/31.
 P-PSDB; B03121.
 New immunogenic recombinant hybrid M protein comprising amino-terminal
 peptide fragments of streptococcal M protein useful as vaccine against
 rheumatic fever and infections leading to rheumatic fever
 Disclosure; Fig 11A-C; 62pp; English.
 The invention relates to multivalent immunogenic hybrid group A
 streptococcal M proteins comprising N-terminal peptide fragments of M
 proteins that elicit opsonic antibodies against multiple serotypes
 of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 generated using the hybrid proteins are against one or more M protein
 serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences A57893-A57902
 CC represent DNAs encoding multivalent hybrid streptococcus pyogenes M
 CC proteins generated in the disclosure of the invention.
 XX
 SQ Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 26.1%; Score 301.8; DB 21; Length 1029;
 Best Local Similarity 68.3%; Pred. No. 3.1e-57;
 Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;
 QY 486 gagagtcgttatactagcgcacgacgaagaatgataaaataattattgacgactct 545
 DB 3 gagagtcgttatactagcgcacgacgaagaatgataaaataattattgacgactct 62
 QY 546 tgacgcaaaagacatgaattacaacacagatgagaagttatctctgcagaacggtga 605
 DB 63 tgacgcaaaagacatgaattacaacacagatgagaagttatctctgcagaacggtgt 122
 QY 606 tggtaatccttaggggaagttatagaagatcttcgacgaacaacatcccgacatacaaat 665
 DB 123 tcttaggggacggtgagaacacccg-----acaaagc 155
 QY 666 acgtttacgtcacgaaacacagagacttaaaagcgagatgagaatgcaatggaagtgc 725
 DB 156 acgagaacactttaaacaagtagcgtagagaactctatgttacaagctaaatgaacaa 215
 QY 726 aggaagagatttttaagagagcgtgcttactgttagatcaggttacacaattatatactaa 785
 DB 216 cttagtcgacgcggtgactaggggtacataataatgacccgcaaaagcgcaaaagagctct 275
 QY 786 acataatagtaattaccacaataataatgcacaagcgtggcagacttgacctgagacaaa 845
 DB 276 tgacaagtagtagctagaaaacccatgacttaaaactaagaatgaaagggttaaaaactga 335
 QY 846 ggctgaatatctaaagccttaataatgattgggctgagagcgtgttacaagagttataatat 905
 DB 336 gaatgaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaac 395
 QY 906 cga-----tgctgcgactagctcagacagatgactctggaataaagtaacaagaactgc 959
 DB 396 tgagccatgggttcgcgactagctcagacagatgactctggaataaagtaacaagaactgc 455
 QY 960 tgacaagtttgagatagaatacaatacgtttaaaacttaagaatgagcttaacttaagttaa 1019
 DB 456 tgacaagtttgagatagaatacaatacgtttaaaacttaagaatgagcttaacttaagttaa 515
 QY 1020 taataaagcgttaaaagatcataatgagtttaaaacttaagaatgagcttaacttaagttaa 1079
 DB 516 taataaagcgttaaaagatcataatgagtttaaaacttaagaatgagcttaacttaagttaa 575
 QY 1080 gaaactacgtataaaatgataatcactatctgaaaaagctagtaaaattcaagaattaga 1139
 DB 576 gaaactacgtataaaatgataatcactatctgaaaaagctagtaaaattcaagaattaga 635
 QY 1140 ggcacgttaagtaaaagcgtt 1158
 |||||

Query Match 20.8%; Score 241; DB 15; Length 861;
Best Local Similarity 98.0%; Pred. No. 5.6e-44;
Matches 244; Conservative 0; Mismatches 5; Indels

streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19. The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a streptococcal M protein; and a method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting

Tue May 15 07:27:08 2001

opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences A57893-A57902 represent DNAs encoding multivalent hybrid streptococcus pyogenes M proteins generated in the disclosure of the invention.

XX Sequence 861 BP; 331 A; 147 C; 199 G; 184 T; 0 other;

Query Match 20.8%; Score 241; DB 21; Length 861;
Best Local Similarity 98.0%; Pred. No. 5.6e-44;
Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaggtacaagaacgctgctgacaagttt 969
DB 4 gtcgcgactaggtctcagacagatactctggaaaggtacaagaacgctgctgacaagttt 63
QY 970 gagatagaaacaatacgtttaaacttaagaatagtgacttaagttttaataataaagcg 1029
DB 64 gagatagaaacaatacgtttaaacttaagaatagtgacttaagttttaataataaagcg 123
QY 1030 ttaaaagatcataatgatgagtttaactgaagatgagtaagtaagtaagtaagtaagtaag 1089
DB 124 ttaaaagatcataatgatgagtttaactgaagatgagtaagtaagtaagtaagtaagtaag 183
QY 1090 aaaaatgataatacactatctgaaaagcgtagtaaaattcaagaattagagcagcgaag 1149
DB 184 aaaaatgataatacactatctgaaaagcgtagtaaaattcaagaattagagcagcgaag 243
QY 1150 taaaagcgtt 1158
DB 244 gctgatctt 252

RESULT 12
ID Q45220 standard; DNA; 522 BP.
XX Q45220:
XX Q45220:
DT 02-NOV-1994 (first entry)
XX Recombinant M24-M5 divalent hybrid gene.
XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
XX M protein; M24; M5; M6; M19; subunit; tetravalent protein; emm gene;
XX tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;
XX restriction enzyme site; multivalent M protein; immunisation; group A;
XX streptococci; rheumatic fever; rheumatic heart disease; humoral;
XX antibody; heart tissue; antigen; serotype; mucosal; ss.

Synthetic.
XX WO9406421-A.
XX 31-MAR-1994.
XX 15-SEP-1993; 93WO-US08703.
XX 16-SEP-1992; 92US-0945954.
XX (UVTE-) UNIV TENNESSEE RES CORP.
XX Dale JB, Lederer JW;

XX WPI; 1994-118122/14.
DR P-PSDB; R50998.
XX New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes
XX Disclosure; Fig 10; 67pp; English.
XX The sequences given in 045214-23 encode hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These tetravalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC pK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.
XX Sequence 522 BP; 226 A; 68 C; 116 G; 112 T; 0 other;

Query Match 20.7%; Score 239.4; DB 15; Length 522;
Best Local Similarity 97.6%; Pred. No. 1.1e-43;
Matches 243; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaggtacaagaacgctgctgacaagttt 969
DB 4 gtcgcgactaggtctcagacagatactctggaaaggtacaagaacgctgctgacaagttt 63
QY 970 gagatagaaacaatacgtttaaacttaagaatagtgacttaagttttaataataaagcg 1029
DB 64 gagatagaaacaatacgtttaaacttaagaatagtgacttaagttttaataataaagcg 123
QY 1030 ttaaaagatcataatgatgagtttaactgaagatgagtaagtaagtaagtaagtaagtaag 1089
DB 124 ttaaaagatcataatgatgagtttaactgaagatgagtaagtaagtaagtaagtaagtaag 183
QY 1090 aaaaatgataatacactatctgaaaagcgtagtaaaattcaagaattagagcagcgaag 1149
DB 184 aaaaatgataatacactatctgaaaagcgtagtaaaattcaagaattagagcagcgaag 243
QY 1150 taaaagcgtt 1158
DB 244 gctgatctt 252

RESULT 13
ID A57899 standard; DNA; 522 BP.
XX A57899;
XX 10-OCT-2000 (first entry)
XX S. pyogenes hybrid M protein (M24-M5) DNA, SEQ ID NO:14.
XX Multivalent hybrid M protein; group A streptococcus; serotype;
XX immunogenic; sero-specific antibody; streptococcal infection;
XX cross reactivity; vaccine; acute rheumatic fever; ARF;
XX rheumatic heart disease; streptococcal pharyngitis; strep throat;
XX pneumonia; ds.

Db 1 atgtcgcgactaggtctcagacagatactctggaataagtaacagaagtcgcgactagg 60
Qy 67 tttagatagataaaacatactgtaaaactaagtaagtagtacttaagtttaataataa 126
Db 61 tctcagacagatactctggaataagtaacagaagtcgcgactaggctcagacagatact 120
Qy 127 gcgttaaaagatcataatgatgatgtaactgaagagtgtagtaagtctaagaagaacta 186
Db 121 ctggaaaagtcacagaagagtcgcgcgtgactagggttacaataaagtcacccaaga 180
Qy 187 cgtggatccgcgtgactagggttacaataaagtcacccaagaagtcacccaagaagtc 246
Db 181 gcaaaaagacgcgtgactagggttacaataaagtcacccaagaagtcacccaagaagtc 240
Qy 247 gacaagtagtgcagtaaaacacacacacacacacacacacacacacacacacacac 306
Db 241 actagggttacaataaagtcacccaagaagtcacccaagaagtcacccaagaagtc 300
Qy 307 aatgaaggttataaaactgagaagtcacccaagaagtcacccaagaagtcacccaaga 366
Db 301 gggacggtagaac 360
Qy 367 gaggtcgac-----agagtttcttaggggacggttagaataagtcacccaaga 420
Db 361 ccggacaaagcagcagaagtggttcttaggggacggttagaataagtcacccaaga 420
Qy 421 gaacttcttaacaagtagcagtagagagaactctatgttacaaagtcacccaaga 480
Db 421 cca-----tggagagtcgcttatactaggcgcacacacacacacacacacacac 465

RESULT 15
A57896
ID A57896 standard; DNA; 561 BP.
AC A57896:
DT 10-OCT-2000 (first entry)
DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:7.
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX Streptococcus pyogenes.
OS Synthetic.
XX US0603386-A.
PN 16-MAY-2000.
PD 15-SEP-1997; 97US-0937271.
PF 16-SEP-1992; 92US-0945954.
PR (UYTE-) UNIV TENNESSEE RES CORP.
XX Lederer JW, Dale JB;
PI WPI: 2000-364475/31.
XX P-PSDB; B03116.
DR New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX

PS Disclosure; Fig 7A-B; 62pp; English.
XX The invention relates to multivalent immunogenic hybrid group A streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19. The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a multivalent hybrid M protein; and a method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences A57893-A57902 represent DNAs encoding multivalent hybrid Streptococcus pyogenes M proteins generated in the disclosure of the invention.

XX Sequence 561 BP; 217 A; 108 C; 151 G; 85 T; 0 other;
SQ

Query Match 10.7%; Score 124.4; DB 21; Length 561;
Best Local Similarity 55.5%; Pred. No. 1.5e-18;
Matches 295; Conservative 0; Mismatches 216; Indels 21; Gaps 2;

Qy 7 atgtcgcgactaggtctcagacagatactctggaataagtaacagaagtcgcgactagg 66
Db 1 atgtcgcgactaggtctcagacagatactctggaataagtaacagaagtcgcgactagg 60
Qy 67 tttagatagataaaacatactgtaaaactaagtaagtagtacttaagtttaataataa 126
Db 61 tctcagacagatactctggaataagtaacagaagtcgcgactaggctcagacagatact 120
Qy 127 gcgttaaaagatcataatgatgatgtaactgaagagtgtagtaagtctaagaagaacta 186
Db 121 ctggaaaagtcacagaagagtcgcgcgtgactagggttacaataaagtcacccaaga 180
Qy 187 cgtggatccgcgtgactagggttacaataaagtcacccaagaagtcacccaagaagtc 246
Db 181 gcaaaaagacgcgtgactagggttacaataaagtcacccaagaagtcacccaagaagtc 240
Qy 247 gacaagtagcagtaaaacacacacacacacacacacacacacacacacacacac 306
Db 241 actagggttacaataaagtcacccaagaagtcacccaagaagtcacccaagaagtc 300
Qy 307 aatgaaggttataaaactgagaagtcacccaagaagtcacccaagaagtcacccaaga 366
Db 301 gggacggtagaac 360
Qy 367 gaggtcgac-----agagtttcttaggggacggttagaataagtcacccaaga 420
Db 361 ccggacaaagcagcagaagtggttcttaggggacggttagaataagtcacccaaga 420
Qy 421 gaacttcttaacaagtagcagtagagagaactctatgttacaaagtcacccaaga 480
Db 421 cca-----tggagagtcgcttatactaggcgcacacacacacacacacacacac 465

Qy 481 ccatggagagtcggttatacttagtcacacacacacacacacacacacacacacacac 532
Db 466 aaaaaagagtcgcttatacttagtcacacacacacacacacacacacacacacacac 517

Tue May 15 07:27:08 2001

Search completed: May 13, 2001, 04:57:44
Job time: 8144 sec

us-09-151-409-15.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:30:06 ; Search time 66.83 Seconds
(without alignments)
3025.436 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcatgcatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	419	36.2	741	3	US-08-937-271-3
2	328.2	28.3	822	3	US-08-937-271-5
3	301.8	26.1	741	3	US-08-937-271-12
4	301.8	26.1	1029	3	US-08-937-271-16
5	241	20.8	861	3	US-08-937-271-1
6	239.4	20.7	522	3	US-08-937-271-14
7	151	13.0	1332	2	US-08-795-475-5
8	124.4	10.7	561	3	US-08-937-271-7
9	82.6	7.1	7218	1	US-08-232-463-14
10	76.2	6.6	408	3	US-08-937-271-21
11	53.2	4.6	204	3	US-08-937-271-19
12	53.2	4.6	918	3	US-08-937-271-9
13	42.6	3.7	6243	2	US-09-056-075-1
14	42	3.6	2223	3	US-08-257-073-4
15	41.6	3.6	6755	3	US-08-931-999-4
16	41.4	3.6	1498	1	US-08-118-469A-1
17	41.4	3.6	1498	1	US-08-909-119-1
18	40.2	3.5	4766	5	PCT-US93-07261-10
19	39.8	3.4	5181	1	US-08-257-073-10
20	39.6	3.4	243	1	US-08-182-175A-56
21	39.6	3.4	243	1	US-08-474-633A-74
22	39.6	3.4	243	5	PCT-US92-06412-56
23	39.4	3.4	5361	4	US-08-973-462-2
24	39.4	3.4	6152	4	US-08-973-462-1
25	38.8	3.4	3279	4	US-08-446-137B-1
26	38.6	3.3	1519	1	US-07-971-759-19
27	38.6	3.3	2447	2	US-09-014-969-14

28	38.6	3.3	3645	2	US-08-663-112-1	Sequence 1, Appli
29	38.4	3.3	453	3	US-08-714-918-9	Sequence 9, Appli
30	38.4	3.3	453	4	US-09-265-315-9	Sequence 9, Appli
31	38.4	3.3	453	4	US-09-265-315-9	Sequence 9, Appli
32	38.2	3.3	2258	1	US-07-720-589-1	Sequence 1, Appli
33	38.2	3.3	2258	2	US-08-785-190-1	Sequence 1, Appli
34	38.2	3.3	2258	5	PCT-US92-05539-1	Sequence 5, Appli
35	38.2	3.3	2672	4	US-09-214-564A-5	Sequence 1, Appli
36	38.2	3.3	2815	4	US-09-214-564A-1	Sequence 1, Appli
37	38.2	3.3	3143	1	US-08-485-621-1	Sequence 1, Appli
38	38.2	3.3	3143	2	US-08-973-831-1	Sequence 1, Appli
39	38.2	3.3	3143	5	PCT-US96-09530A-1	Sequence 1, Appli
40	38	3.3	39	3	US-08-937-271-26	Sequence 1, Appli
41	38	3.3	47	3	US-08-937-271-28	Sequence 1, Appli
42	38	3.3	1320	1	US-08-257-073-15	Sequence 26, Appli
43	37.8	3.3	5433	4	US-08-929-329-1	Sequence 15, Appli
44	37.8	3.3	19124	2	US-08-487-826B-13	Sequence 13, Appli
45	37.6	3.2	1560	1	US-07-813-584A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-937-271-3
; Sequence 3, Application US/08937271
; Patent No. 6053386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
US-08-937-271-3

GENERAL INFORMATION:
APPLICANT: Date, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

RESULT 3
US-08-937-271-12

Sequence 12, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..741
US-08-937-271-12

Query Match 26.1%; Score 301.8; DB 3; Length 741;
Best Local Similarity 68.3%; Pred. No. 1.4e-68;
Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;
QY 486 gagagtgcttactagccatgcgcagagataagctaaataattattgacgatct 545
|||||
Db 3 GAGAGTCGTTACTAGGATAGCCAGAGATAGCTAAAAAATAATTATGACGATCT 62
QY 546 tgacgaaaagacatgaattacaacagagatgagagattctctgcagacggtga 605
|||||
Db 63 TGACGCAAAAGACATGAATTAACAACAGAGATGAGAGTTATCTGGATCCAGAGTGT 122
QY 606 tggtaacctaggaagattagaagaattctgcagcaaacatcccgcaatacaaaatat 665
|||||
Db 123 TCCTAGGGGACGGTAGAATCCCGG-----ACAAAGC 155
QY 666 acgtttactgcagaaaacagacttaaacgcgagattagagaatgcaatggaagtgc 725
|||||
Db 156 ACAGAACTTCTTAAACAGTATGACGTAGAACCTATGTTACAGACTTAATAATGACAA 215
QY 726 aggaagagatttaagagagctggtacctgttagatcaggtgtacacattatatactaa 785
|||||
Db 216 CTTAGTCGCGCGTACGATAGGGGTACAATAATGACCGCAAGAGCAAGAGAGCTCT 275
QY 786 acataatagtaattaccacaataataatgcacaaagctggcagacttgacctgagacaaa 845

Db 276 TGACAAGTATGAGCTAGAAAACCATGACTTAAAAACTAAGAAATGAAGGTTAAAAACTGA 335
QY 846 ggctgaatatctaaagcccttaattgctggcgtgagagctgtttacaagagttaaatat 905
|||||
Db 336 GAATCAAGGGTTAAAAACTGAGATGAAGGTTAAAAACTGAGAAATGAAGGTTAAAAAAC 395
QY 906 cga-----tgtcgactaggtctcagacagatctctgaaaaagctacaagaacgtgc 959
|||||
Db 396 TGAGCCATGGTGCAGCTAGGTCTCAGACAGATACTCTGAAAAAGCTACAGAACGTGC 455
QY 960 tgacaagtttgagatagaaaaacatactttaaaacttaagaacttaagaacttaagtttaa 1019
|||||
Db 456 TGACAAGTTTGAGATAGAAAACATACTTAAAGCTTAAAGATAGTGAAGTTAAAGTTTAA 515
QY 1020 taataagcgttaaaagatcatatgatgatgattaaactgaagagttgagtaactgaaga 1079
Db 516 TAATAAAGCGTTAAAGATCATATGATGATGATTAACCTGAAGAGTTGAGTATGCTTAAGA 575
QY 1080 gaaactacgtaaaaatgataaaatcaactatctgaaaaagctagtaaaattcagaattaga 1139
Db 576 GAACTACGTAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAAAATCAAGAAATTAGA 635
QY 1140 ggcacgtgaagtaaaagctt 1158
|||||
Db 636 GGCACGTAAAGCTGAICTT 654

RESULT 4
US-08-937-271-16
Sequence 16, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1029 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 1..1029
US-08-937-271-16

Query Match      26.1%; Score 301.8; DB 3; Length 1029;
Best Local Similarity 68.3%; Pred. No. 1.5e-68;
Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtgcttatactaggcgcacgcagagataaagctaaagaaattattgacgatct 545
DB 3 GAGAGTGGCTTACTAGGCATACGCCAGAGATAAGCTAAAAAATAATTATGACGATCT 62
QY 546 tgacccaaagaaacatgaattacaaacagagaatgagaattctctcgcagaaagtgga 605
DB 63 TGACGCAAAAGAACATGAATTAATACAAACAGCAATGAGAAGTTATCTGGATCCAGAGTGT 122
QY 606 tggtaactcctagggaagttatagaagatcttgccagcaacaatcccgcaatacaaaatat 665
DB 123 TCCTAGGGGGCGGTAGAAACCCGG-----ACAAAGC 155
QY 666 acgtttacgtcagcaaaacagagcttaaaagcgagattgagagaatcgaatggaagtgc 725
DB 156 ACGAGAACTTCTTAAACAAGTATGAGTAGAGAACTCTATGTACAGACTAATAATGACAA 215
QY 726 aggaagagatttaagagagctggtaccttgttagatcaggttacacaattatatactaa 785
DB 216 CTTAGTCGACGCCGTGACTAGGGGTGCAATAAATGACCCGCAAGAGCAAAAGAGCTCT 275
QY 786 acataatagtaattaccacaataataatgcacaagctggcgagacttgacctggcgacaaa 845
DB 276 TGACAAGTATGAGCTAGAAAACCATGACTTAAAAACTAAGAATGAAGGTTTAAAAACTGA 335
QY 846 ggcgaatctctaaagagccttaagtattggtgagagctgttacaaaggttaaatat 905
DB 336 GAATGAAGGTTTAAAAACTGAGAATGAAGGGTTAAAAACTGAGAAATGAGGGTTAAAAAC 395
QY 906 cga-----tgtcgactaggtctcagacagatactctggaaaaaagtacaagaacgtgc 959
DB 396 TGAGCCATGGTTCGGCAGCTAGTCTCAGACAGATACTCTGGAAAAAGTACAAAGAACGTGC 455
QY 960 tgacaagtttgagatgaaacaaacacgtttaaacttaagaatagtgacttaagtatttaa 1019
DB 456 TGACAAGTTTGAGATGAGAAACAATACGTAAAACTTAAGAAATAGTGACTTAAGTTTAA 515
QY 1020 taataagcgttaaaagatcataatgatagtttaactgaaggttgagtaagtgcataaga 1079
DB 516 TATTAAGCGTTTAAAGATCATATGATGAGTTAACTGAAGAGTTGAGTAATGCTAAAGA 575
QY 1080 gaaactacgtataaaatgataatcactatctgaaaaaagctagcaaaattcaagaattaga 1139
DB 576 GAAACTACGTAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAAATCAAGATTAGA 635
QY 1140 ggaacgttaagtaaaagcctt 1158
DB 636 GGCACGTAAGGCTGATCTT 654

RESULT 5
US-08-937-271-1
; Sequence 1, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

; NAME/KEY: CDS
; LOCATION: 1..861
US-08-937-271-1

Query Match      20.8%; Score 241; DB 3; Length 861;
Best Local Similarity 98.0%; Pred. No. 5.2e-53;
Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 910 gtgcgactaggtctcagacagatactctgaaaaaagtacaagaacgtgctgacaagttt 969
DB 4 GTCCGACTAGGTCTCAGACAGATACTCTGAAAAAAGTACAAAGACGTGCTGACAAAGTTT 63
QY 970 gagatagaaacaatacagcttaaaacttaagaatagtgacttaagttttaataataaagcg 1029
DB 64 GAGATAGAAAACAATACGTTAAACTTAAGATAGTAGTAACTTTAATTAATAAAGCG 123
QY 1030 ttaaaagatcataatgatgatgagtttaactgaagagttgagtaatgctaaagaaactacgt 1089
DB 124 TTAAGAAGATCATAATGATGAGTTAACTGAAGAGTTGAGTAATGCTAAAGAGAACTACGT 183
QY 1090 aaaaatgataaatcactatctgaaaaagctagtaaaattcaagaattagagcgcaag 1149
DB 184 AAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAANTTCAGAAATTTAGAGCGCAGTAAG 243
QY 1150 taaaagcctt 1158
DB 244 GCTGATCTT 252

RESULT 6
US-08-937-271-14
; Sequence 14, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
```



```
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..519
; US-08-937-271-14
```

```
Query Match      20.7%; Score 239.4; DB 3; Length 522;
Best Local Similarity 97.6%; Pred. No. 1.1e-52;
Matches 243; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 910 gtcggaactagctcagacagatctctggaagaaagtcacaaagctgctgacaagttt 969
Db 4 gtcggagactagctctcagacagatctctggaagaaagtcacaaagctgctgacaagttt 63

Qy 970 gagatagaaacaatagcttaaaacttaagaatagtgacttaagttttaataaagaagc 1029
Db 64 GAGATAGAAAACATACGTTAAACCTTAAAGATAGTGAAGTTTAAATATAAAGCG 123

Qy 1030 taaaagatcataatgatgatgacttaactgaagagtgtagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGAATCATATAATGATGAGTTAACTGAAGAGTTGAGTAATCTAAAGAGAACTACGT 183

Qy 1090 aaaaatgataatcactatctgaaaaagctagtaaaattcaagaattagaggcacgtaag 1149
Db 184 AAAAATGTAATCACTATCTGAAAAGCTAGTAAAAATCAAGAAATAGAGGCACGTAAG 243

Qy 1150 taaagctt 1158
Db 244 GCTGATCTT 252
```

```
RESULT 7
US-08-795-475-5
; Sequence 5, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
```

```
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1329
; US-08-795-475-5
```

```
Query Match      13.0%; Score 151; DB 2; Length 1332;
Best Local Similarity 100.0%; Pred. No. 6.4e-30;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 598 aacgggtgatggaactctctagggagttatagaagatcttgacgaaacaatcccgcaata 657
Db 1 AACGGTGATGTTAATCTCTAGGGAGTTTATAGAAGATCTTCAGCAAAACAATCCCGCAATA 60

Qy 658 caaataatcgtttacgtcacgaaacaagacttaaaagcgagattagagaatgcattg 717
Db 61 CAAATATACGTTTACGTCACGAAACAGGACTTAAAGCGAGATTAGAGATGCAATG 120

Qy 718 gaagttcaggaaagagattttaagagagctg 748
Db 121 GAAGTTCAGGAAGAGATTTTAAAGAGAGCTG 151
```

```
RESULT 8
US-08-937-271-7
; Sequence 7, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-937-271-7

Query Match 10.7%; Score 124.4; DB 3; Length 561;
Best Local Similarity 55.5%; Pred. No. 3e-23;
Matches 295; Conservative 0; Mismatches 216; Indels 21; Gaps 2;
QY 7 atgttcgactgagttctcagacagatactctggaagaaagtaagaacgctgtaagaag 66
DB 1 ATGTCGCGACTAGGTCTCAGACAGATACCTCTGGAAAAAGTACAAGAGTCGCGACTAGG 60
QY 67 tttagatagaatacaatacgtttaaaacttaagaatagtgacttaagttttaataataa 126
DB 61 TCTCAGACAGATACCTCGGAAAAAGTACAAGAGTCGCGACTAGGTCTCAGACAGATAC 120
QY 127 gcgttaaaagatcataatgatgatgagtttaactgaagagttgagtaagtctaaagaaacta 186
DB 121 CTGAAAAAAGTACAGAAGAGTCCGCCGTGACTAGGGGTACAATAAATGACCCGCAAGA 180
QY 187 cgtggtatccgctgactgaggtgacataataatgacccgcaagagcaaaagaagctctt 246
DB 181 GCAAAAGAACCGGTGACTAGGGGTACANTAAATGACCCGCAAGAGCAAAAGAACCGGTG 240
QY 247 gacagtatgagctagaaaaaccatgacttaaaactaagaatgaagggttaaaactgag 306
DB 241 ACTAGGGGTACAATAAATGACCCGCAAGAGCAAAAGAGTCGACAGAGTGTTCCTAGG 300
QY 307 aatgaaggggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaact 366
DB 301 GGGACGGTAGAAAACCCGGACAAAGACGACGAGAGTGTTCCTAGGGGGACGGTAGA 360
QY 367 gaggctgcac-----agagtggtttcctaggggacggtagaaaacccggacaaagcaga 420
DB 361 CCGGACAAAGCACGAGAGAGTGTTCCTAGGGGGACGGTAGAAGAACCCGCAAGACGA 420
QY 421 gaactcttaacaagatgacgtgagaaactctgtttacaactaataatgacaagtta 480
DB 421 CCA-----TGAGAGTGGCTTATACCTAGGCATACGCCAGAGATAGCTA 465
QY 481 ccattgagagtgctgtatactaggcactacgcccagaagataaagcttaaaaaaa 532
DB 466 AAAAAAGAGTGGCTTATACCTAGGCATACGCCAGAGATAGCTAAAAAAA 517

RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463

Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 7.1%; Score 82.6; DB 1; Length 7218;
Best Local Similarity 8.9%; Pred. No. 3.7e-12;
Matches 40; Conservative 239; Mismatches 168; Indels 0; Gaps 0;
QY 115 tttaataaaacgcttaaaagatcataatgatgagtttaactgaagagttgagtaagt 174
DB 1511 TTTCAAAAACGGCATGTAGGCATCACTGTAATCTATCTATGCAAGTACTTAAGAG 1452
QY 175 aaagagaaactcgtgacgtccgctgactaggggtacataataatgacccgcaagagca 234
DB 1451 ATACAAGATTGCTACRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1392
QY 235 aaagaaactcttgacaagtatgactgagaaacacacactgacttaaaactaagaatgaagg 294
DB 1391 RNR 1332
QY 295 ttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactgagaatgaagg 354
DB 1331 RNR 1272
QY 355 ggggttaaaactgaggggtcagacagagtttccctaggggagcgttagaaaacccggacaaa 414
DB 1271 RNR 1212
QY 415 gcacgagaaactcttcaacaagtatgacgtagagaactctatgttaacagctaataatgac 474

APPLICANT: Dale, James B.
 APPLICANT: Lederer, James W.
 TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
 TITLE OF INVENTION: VACCINE
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release 1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/937,271
 FILING DATE: 15-SEP-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Rosenman, Stephen J.
 REGISTRATION NUMBER: 43,058
 REFERENCE/DOCKET NUMBER: 481112.405C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 918 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus pyogenes
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..915
 OS-08-937-271-9

Query Match	4.6%	Score 53.2;	DB 3;	Length 918;
Best Local Similarity	87.9%;	Pred. No. 5.9e-05;		
Matches 58;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0;
Qy	176	aagagaactacgttgatcccccgtgactagggggtacaataaatgatgccgcgaagagcaa	235	
Db	35	AAAAAGTACAAGAAGGATCCCGCTGACTAGGGGTACATAAATGATCCGCCAAAGAGCAA	94	
Qy	236	asgaag	241	
Db	95	AAGAAG	100	

RESULT 13
 US-09-056-075-1
 ; Sequence 1, Application US/09056075
 ; Patent No. 5955368
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Eric A.
 ; APPLICANT: Bradshaw, Marite
 ; APPLICANT: Rood, Julian
 ; TITLE OF INVENTION: Expression System for Clostridium
 ; TITLE OF INVENTION: Species
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Pinckney Street
 ; CITY: Madison

```

STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,075
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95238
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3770..4013
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (oriT) from
OTHER INFORMATION: plasmid RP4"
US-09-056-075-1

```

	Query Match	3.7%	Score 42.6;	DB 2;	Length 6243;
	Best Local Similarity	53.3%;	Pred. No. 0.06;		
	Matches	90;	Conservative	0; Mismatches 79; Indels 0; Gaps	
Qy	971	agatgagaaacaatc	ggttaaa	aacttaagaatagtgacttaagttttaataataaagcgt	1030
Db	1217	AAATAAAAAAAT	AAAAAAT	AAAAAAT	AAAAAAT
Qy	1031	taaaagatcatcatgatgatg	ctcaactgaaggttgagtaactgc	caagagagaacactagctg	1090
Db	1277	TAAAAAAT	TAAAAAAT	TAAAAAAT	TAAAAAAT
Qy	1091	aaaatgat	aaatcactatctg	aaaaagctagtaaattccaagaattaga	1139
Db	1337	AAAT	AAAAAAT	AAAAAAT	AAAAAAT

```

RESULT 14
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Time, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-4
```

Query Match 3.6%; Score 42; DB 1; Length 2223;
Best Local Similarity 46.3%; Pred. No. 0.06; Indels 0; Mismatches 160; Gaps 0;

```
QY 83 atactttaaacttaagaatagtgtaattttaaataaaagcggttaaagaatcata 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1892 AAATATTTCAACAAAATATTCTAGAAAATGATGTTCTTAATCAAGAAACGGAGGAAGAA 1951

QY 143 atgatgagtgtaactgaagtgtaagtctaaagagaaactacgtggatccgccgtga 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1952 TGGAAAACACAGTGTGACANTACCAAGCAAAATAGAAAGTGAAGTGGATGCCCTCGCAC 2011

QY 203 ctagggtgtacataaatgaccgcgaagagcaaaagcgctcttgacaagtatagctag 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2012 CAAAATAATAGGAAGAGAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 2071

QY 263 aaaaactgacttaaaactaagaatgaagggtttaaaactgagaatgaagggttaaaaa 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2072 AAAAGAGAGAAAAGAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 2131

QY 323 ctgagaatgaagggttaaaaactgagaatgaagggtttaaaactgagtcgacagagt 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2132 AAGAAGAGAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGT 2189
```

```
RESULT 15
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007
; US-08-931-999-4
```

Query Match 3.6%; Score 41.6; DB 3; Length 6755;
Best Local Similarity 47.5%; Pred. No. 0.11;
Matches 134; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

```
QY 42 aaagtacaaagacgtgctgacaaagtttgagatagaaaacaatacgtttaaacttaaga 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6112 AACCGAAAAAGCACGGAGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6171

QY 102 tagtgacttaagttttaataataaaagcgtttaaagatcataatgatgatgaagaaga 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6172 AACGGGACAGAGACGACCAAGAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6231

QY 162 gttgagtaagtctaaagagaaactacgtgactccgcgtgactaggggtacataataatga 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6232 AGAGAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 6291

QY 222 ccgcgaagagcaaaagaagctcttga-caagtgatgagctagaacacccatgacttaaaaa 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6292 AACCCGCAAGAGCAGGCCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6351

QY 281 ctaagaatgaagggtttaaaactgagaatgaagggtttaaaactgagaatgaagggttaa 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6352 CCAAAAACAAAAGGAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6411

QY 341 aaactgagaatgaagggttaaaaa 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6412 ACAGAGAGAAACCAAGCCCAAAAAA 6435
```

Search completed: May 13, 2001, 04:59:08
Job time: 1742 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 01:57:50 ; Search time 1036.68 Seconds
(without alignments)
9758.461 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatgctgcgactag.....aggcacgtaagtaaaagcgtt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query		DB	ID	Description
			Match	%			
C	1	74.8	6.5	997	229	CNS005TE	AL060767 Drosophil
	2	72.4	6.3	839	229	CNS0122R	AL101037 Drosophil
C	3	71.8	6.2	1447	146	BF273312	BF273312 GA_EB001
	4	70.6	6.1	1885	166	BE420745	BE420745 HWM002.B0
C	5	69	6.0	1282	146	BF264918	BF264918 HV_CEA001
	6	68	5.9	1036	231	CNS03LWJ	AL250012 Tetraodon
	7	67.2	5.8	1106	146	BF264948	BF264948 HV_CEA001
C	8	66.4	5.7	1223	227	BI2981	BI2981 T24D11-Sp6
	9	66	5.7	907	230	CNS021J4	AL176953 Tetraodon
C	10	65.4	5.6	1101	229	CNS0039G	AL063921 Drosophil
	11	65	5.6	1101	229	CNS017RP	AL108415 Drosophil
	12	64.8	5.6	853	223	AZ545275	AZ545275 ENTGB15TF
C	13	64.8	5.6	899	223	AZ533253	AZ533253 ENTBI02TF
	14	64.6	5.6	1036	229	CNS00599	AL057797 Drosophil
C	15	64.2	5.5	1101	229	CNS0182P	AL108811 Drosophil
	16	64.2	5.5	1135	230	CNS033GQ	AL226115 Tetraodon
C	17	63.8	5.5	1101	229	CNS0039L	AL063926 Drosophil
C	18	63.8	5.5	1184	227	BI3117	BI3117 T9K3-Sp6.2


```
QY 1136 tagaggcagtaagtaaaagctt 1158
Db 28 TNNNNNNNNNNNNNTATNWT 6

RESULT 2
LOCUS CNS0122R 839 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN07E20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL101037
VERSION AL101037.1 GI:5612648
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmod Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 839)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES
Source Location/Qualifiers
1..839
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN07E20"
/note="end : SP6"

BASE COUNT 511 a 54 c 107 g 85 t 82 others
ORIGIN

Query Match 6.3%; Score 72.4; DB 229; Length 839;
Best Local Similarity 41.9%; Pred. No. 1.1e-06;
Matches 226; Conservative 38; Mismatches 274; Indels 2; Gaps 1;

QY 40 gaaaagtaagaagcgtctgacagtttgagatagagaacaatacgttaaaacttaag 99
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 76 GTATAAAACMTGAAAGTAGAATATATWTTWGCAGGAAAAAAAMWTTAAAGCACAWAAA 135

QY 100 aatagtgacttaagttttaataaagcgttaaaagatcataatgagtgactaa 159
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 136 AAAAAAAAMWAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAA 195

QY 160 gaggtagtaagtaagaagaaactcgtggatccgcgtgactaggggtacaaataat 219
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 196 AAGAACAAAAAARAAAGAGAGMAAAAAAARAAAAAAGAAAAAAGAAAAAAGAAAA 255

QY 220 gaccgcgaagagcaaaagacgtcttgcaagtgatgagctagaaacaccatgacttaaa 279
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 256 RAAAAAAMWCGAARGGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 315

QY 280 actaagaatg--aagggttaaaactgagaatgaagggttaaaactgagaatgaagggt 337
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 316 AAAAAAAMWAAAGGAARAAAAAARAAACCAAAAAAARAAAAAAGAAAAAAGAAAGA 375

QY 338 taaaaactgagaatgaagggttaaaactgaggtcgacagagtggttctctaggggaagg 397
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 376 AAAAAAAGCAAAAMAGGACAAAAAARGAACRCGGAARAAAAAARAAAAAAGAAAGCA 435
```

```
QY 398 tagaaacccggacaaagcacgcgaactcttaacaagtatgacgtagagaactctatgt 457
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 436 ARCAMACARAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 495

QY 458 tacaagctaataatgacaagttaccatggagagtgcggttatactaggcaccagaag 517
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 496 AAGAARAAAAAMGGAAAAAARAAARAAARAAARAAARAAARAAARAAARAAARAA 555

QY 518 ataagctaaaaaattattgacgactcttgacgcacaaagaaacatgaattacaacaacaga 577
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 556 AAAAAAAMWAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 615

RESULT 3
BF273312/c 1447 bp mRNA EST 17-NOV-2000
LOCUS GA_Eb0017107f Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboreum cDNA clone GA_Eb0017107f, mRNA sequence.
ACCESSION BF273312
VERSION BF273312.1 GI:11204382
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 1447)
AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATAGCACTCACTATAGGG
High quality sequence start: 489
High quality sequence stop: 930.

FEATURES
Source Location/Qualifiers
1..1447
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Eb0017107f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 91 a 45 c 9 g 724 t 578 others
ORIGIN

Query Match 6.2%; Score 71.8; DB 146; Length 1447;
Best Local Similarity 38.5%; Pred. No. 1.6e-06;
Matches 346; Conservative 0; Mismatches 552; Indels 0; Gaps 0;

QY 257 agctagaaaaaccatgacttaaaactaagaatgaagggttaaaactgagaatgaagggt 316
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1411 AGAAGAAAAAAGAGAGAGGAAAAAGATAAAAAAGTAAAAAGAAAAAAGAAAAA 1352

QY 317 taaaactgagaatgaagggttaaaactgagaatgaagggttaaaactgaggtcgaca 376
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1351 AARAGAGAAAAAATATATAAAAAGAAAAAATAAAAAATAAAAAATATATAA 1292

QY 377 gagtgttctctaggggacgttagaaaaaccgagcaagcagcagaactcttcaacaagt 436
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1291 TAAAAAAMWAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 1232
```

QY 437 atgacgtagagaactctatgtttacaaagctaataatgacaagttaccatggagagtgogtt 496
 Db 1231 TCAAGGTATATAAAATAAATAATATAAATAAATAATATAAATAAATAAATAAATAA 1172
 QY 497 atactaggcatccgcagagaataagcttaaaaaaattattgacgtatcttgacgcacaaag 556
 Db 1171 AAATTATATAAATAATATAATGAGAAAAAAGAGAAAAAAGAAAAAATAAATAAATAA 1112
 QY 557 aactgtaattacacacagaatgagaagttatctctgcagaacggtgagtgtaactcta 616
 Db 1111 AAAGGAAAGAAAGAGAGTAGTAAATTTTAAANNAANNAAGATNAATAAATAAATAA 1052
 QY 617 gggaggttagaagattcttcagcaacaactccgcacaatacaataatagctttacgtc 676
 Db 1051 AATTAATAAGATTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 992
 QY 677 acgaaacacaggacttaaaacgagatttagagaatgcaatggaagttgcaggaagagatt 736
 Db 991 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 932
 QY 737 ttaagagagctgggtaccttggtagatcagggttacacaattatatactaaacataatagta 796
 Db 931 TTAAGGNAATAAATATNNNTAANNNTGGAATAATAAAGATAAATAAATAAATAAAG 872
 QY 797 attacaacaataatgacagaactggcagacttgacctgagacaaaaggcgaatc 856
 Db 871 AAAAAAATAAATAAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 812
 QY 857 taaagccctaatgctgggctgagaggtctgttacaagagtttaaatcagatgctgcga 916
 Db 811 AANANNNANNAANANNNANNNANNAANNAATTAANNANNNANNAANNAANNAANNA 752
 QY 917 ctggtctcagacagatactctggaagaagtcagaagcgtgctgacaggtttgagatag 976
 Db 751 NNANNNANNNANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 692
 QY 977 aaaaacacgtttaaaccttaagatactgacitaaagtttataataaagcgttaaaag 1036
 Db 691 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 632
 QY 1037 atcataatgtagtgaactgagaggttagtgaatgctaaagagaactacgtataaatg 1096
 Db 631 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 572
 QY 1097 ataatcactatctgaaagcgtagtaaatcagaagtttagaggcagcgtgaagtaaa 1154
 Db 571 ANAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 514

RESULT 4
 BE420745 1885 bp mRNA EST 24-JUL-2000
 LOCUS BE420745 1885 bp mRNA EST 24-JUL-2000
 DEFINITION HW002.B02 ITSC HMW Barley Leaf Library Hordeum vulgare cDNA clone
 ACCESSION BE420745
 VERSION HW002.B02, mRNA sequence.
 KEYWORDS BE420745.1 GI:9418588
 SOURCE EST.
 ORGANISM barley.
 Hordeum vulgare
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
 Hordeum
 1 (bases 1 to 1885)
 Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
 Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
 Pecchioni,N., Qaluset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
 Sorrells,M., Warburton,M. and Wenzel,G.
 International Triticeae EST Cooperative (ITREC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)

COMMENT Contact: Herrmann RG
 Botanisches Institut der LMU
 Menzinger Str. 67, D-80638 Munchen GERMANY
 Fax: 49 30 171683
 Email: herrmann@botanik.biologie.uni-muenchen.de
 International Triticeae EST Cooperative (ITREC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 FEATURES
 source
 1..1885
 /organism="Hordeum vulgare"
 /cultivar="Bake"
 /db_xref="taxon:4513"
 /clone="HW002.B02"
 /clone_lib="ITREC HMW Barley Leaf Library"
 /tissue_type="leaf"
 /dev_stage="14 day old"
 /note="vector: pBluescriptSK(-); 850 bp average insert size."
 BASE COUNT 1138 a 219 c 212 g 176 t 140 others
 ORIGIN
 Query Match 6.1% Score 70.6; DB 166; Length 1885;
 Best Local Similarity 37.5%; Pred. No. 3.2e-06;
 Matches 370; Conservative 0; Mismatches 617; Indels 0; Gaps 0;
 QY 153 aactgagaggttgagtaatgctaagagagaactacgtgacgtgcgtgactagggtac 212
 Db 890 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 949
 QY 213 aataatgacccgcagagcaagagcctctgacaagtagtgacgtagcaaaacacatga 272
 Db 950 NNNNNNNNNNNNAANNAANNAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1009
 QY 273 cttaaaacctaagatgaaggtttaaaactgagaaggttaaaactgagaaggttaaaactgaga 332
 Db 1010 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1069
 QY 333 aggggttaaaactgagaaggttgaggggttaaaactgaggtgcgacagagtggttcctaggg 392
 Db 1070 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1129
 QY 393 gacggttagaaacccgcagacagcagacacttcttaacagtagtgacgtagagaactc 452
 Db 1130 NNAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1189
 QY 453 tatgttacaagctaataatgacaagttaccatggaggtgcgttatactacgtacgcc 512
 Db 1190 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1249
 QY 513 agaagataagcttaaaaaaattatgacgtcttgacgtctgcgcgcaaaagacatgattacaaca 572
 Db 1250 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1309
 QY 573 acagaatgagaaggttatctctgcagaacggtgtagtgaactcctcctagggaggttagaaga 632
 Db 1310 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1369
 QY 633 tcttcagcaacaactccgcgaataataatactattacgtcgcgcaaaagacaggtt 692
 Db 1370 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1429
 QY 693 aaaaacgagattagagaatgcaatggaagttgcaggaagagattttaagagaggtggtac 752
 Db 1430 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1489
 QY 753 ctgttagatcaggttacacattataatacaataataatagtaataaccacataataa 812
 Db 1490 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1549
 QY 813 tgcagaagctggcagactgacccgtgagacaaaggcgtgaatctaaaggccttaaga 872
 Db 1550 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1609

[illegible][illegible][illegible]

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
BF264918	BF264918.1	GI:11195912	EST			
			barley			
			Hordeum vulgare			
			Eukaryota; Viridiplantae;			
			Magnoliophyta; Liliopsida; Poales;			
			Hordeum			
			1 (bases 1 to 1282)			
			Wing, R., Close, T. J., Klein			
			Anderson, H., Dale, J.,			
			T., Saski, C., Schwartzbeck,			
			Wood, T.			

TITLE Development of a genetically and physically anchored
JOURNAL for barley genomics
COMMENT Unpublished (2000)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel.: 864 656 7286
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: AATTACCTCACTAAGGG
 High quality sequence start: 354
 High quality sequence stop: 806.

```

FEATURES
source
...5: 1 Location/Qualifiers
1..1282
/organism="Hordeum vulgare"
/cultivar="Cil16155 (Mla13)"
/db_xref="taxon:4513"
/clone="HV_GEA0010123f"
/clone_lib="Hordeum vulgare seedling green leaf EST"
library HVCN0A0004 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="TJCL121"
/notes="vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
5 a 0 G 524 t 752 others

```

BASE COUNT	6 a	0 c	0 g	524 t	752 others
ORIGIN					
Query Match		6.0%	Score 63;	DB 146;	Length 1282;
Best Local Similarity		27.5%;	Pred. No. 7e-06;		
Matches	192;	Conservative	0;	Mismatches 506;	Indels 0;
					Caps 0;

QY	843	aaaggctgaatatctctaaaggccttaagtattggctgagaggctgttataagagattaaa	902
Db	569	AA	728
QY	903	tatcgatgcgcgactaggtctcagacagatactctggaagaaagtcacgaacgtctga	962
Db	729	AA	788
QY	963	caagtttgagatagaaacaatacgtttaaaacttaagaatagtactagttttaataa	1022
Db	789	AA	848
QY	1023	taaagcgtttaaagatcataatgatgagcttaactgagagtgtagtaatgctctaaagaa	1082
Db	849	AA	908
QY	1083	actacgtgaaatgataaatcactatctgaaagagctagtaaatccaagattagagcc	1142
Db	909	AA	968
QY	1143	acgtgaagtaaaa	1154
Db	969	AAAAAAAAAAAAAA	980
RESULT 7			
BF264948			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			

[illegible]

Query Match 5.8%; Score 67.2; DB 146; Length 1106;
 Best Local Similarity 40.4%; Pred. No. 1.8e-05;
 Matches 333; Conservative 0; Mismatches 491; Indels 0; Gaps 0;

/ 35 ctctggaaaaagtacaagaactgctgcgaagtttgatagatagaacaatacgttaaac 94
 250 CTAANANACACACAAANNCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAA 309
 / 95 ttaagaatagtgacttaagttttaataaagaagcttaaaagatcataatgatgagtaa 154
 310 AAAAAAANNNNNTAAANANAAATAAAAAANNNNNNNNNNNNNNNNNNNNNNNNN 369
 / 155 ctgaagaagttgagtaatgcctaaagagaactacgttgatccgctgactgaggtacaa 214
 370 TAAAAAANTNNNNNNAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 429
 / 215 taatgacccgcgaagagcgaagaagctcttgacaagtatgagttagaaccatgact 274
 430 AAAAAAANNN 489
 / 275 taataactaagaatgaaggggttaaaactgagatgaaggggttaaaactgagatgaag 334
 490 AAAAAAANNN 549
 / 335 ggttaaaactgagatgaatgaaggggttaaaactgaggtcgacagagtggttcctaggggga 394
 550 NAAAAAANNN 609
 / 395 cgtgagaacacccggcacaagacgagaaactcttaacaagtatgactgagaaactcta 454
 610 TAATATTAAATAAATAAATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 669
 / 455 tgttaacagctaataatgacaggttaccatggaggtgcgttatactagtcatacgccag 514
 670 AAAAAAANNN 729
 / 515 aagataagctaaaaaaattattgacgactctgcgcgaagaagaacatgaattacaacac 574
 730 AAAAAAANNN 789
 / 575 agaatgagaagttatctcgcgaagcgtgatcgtaactcctagggaggttatagaagatc 634
 790 TAATATAAATAAANNN 849
 / 635 ttgcagcaacaatccgcgaatacacaatacactgttctgcgcgaagaacagagacttaa 694
 850 TAAAAATTAAATATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 909
 / 695 aagcgaagattagagaatgcaatgaatgctgaggaagatttttaagagagctggacct 754
 910 TATAAATAAATAAANNN 969
 / 755 tgtgagtcaggttcacacaattataactaaacataatagtaattaccacaacataatg 814
 970 AAATATAAANNN 1029
 / 815 cacaagctggcagacttgactgagacaaaggctgaatatcta 858
 1030 AAAAAAANNN 1073

RESULT 8
 B12981/c B12981 DNA GSS 14-MAY-1997
 LOCUS T24D11-Sp6 TBM Arabidopsis thaliana genomic clone T24D11, DNA
 DEFINITION sequence.
 ACCESSION B12981
 VERSION B12981.1 GI:2094103
 KEYWORDS GSS
 SOURCE Chale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Tue May 15 07:27:16 2001

```

Db 789 DDRDAATWTTTITTTTDDDKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKKRRFT 730
Qy 959 ctgcaagtttgagatagaacaataacttaaaacttaagatactgacttaagtta 1018
Db 729 WKRRKRDRDTWDDADADDATARDRRRRGGDDGAGKGTGKRRRRDRATWDRTDAMW 670
Qy 1019 ataataaagcgttaaaagatacataatgatgatgtaactgaagagtgatgactgaag 1078
Db 669 ADAANWTTTDTTDDDKRDRRRKGARRRRRTTARAANDWMTWKAWDKWAKWKTADRW 610
Qy 1079 agaaactactgataaaatgataataactcactatctgaaaaagct 1119
Db 609 DRWAADTWTDAKADRWKAKARARARRRRDRARAARDRWT 569

RESULT 11
LOCUS CNS017RP 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108415 GI:5628719
VERSION AL108415
KEYWORDS fruit fly.
ORGANISM Plasmid Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (BDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelBAC11.

FEATURES             Location/Qualifiers
     source            1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37J10"
                     /note="end : SP6"
BASE COUNT          258 a 174 c 277 g 120 t 272 others
ORIGIN
Query Match          5.6%; Score 65; DB 229; Length 1101;
Best Local Similarity 27.8%; Pred. No. 5.7e-05;
Matches 107; Conservative 104; Mismatches 174; Indels 0; Gaps 0;
Qy 39 gaaagaagtacagaacgtgctgacaagtttgagatagaacaataactgtaaaactaa 98
Db 660 SGRAARASGRCGAGARAARAAGAGAGARRARRARARARARARARARARARAGAGA 719
Qy 99 gaatagtgacttaagtttaataaaacgttaaaagatcataatgatgagtttaactga 158
Db 720 RRRRGAGGAGAGAGAAAAAARARARARARARARARARARARARARARARARARAR 779
Qy 159 agagttgagtaagtctaaagagaactgagtcgcgcgtgactaggggtacataaa 218
Db 780 AGAARGAGAAARRRRRRAGRRGGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 839
Qy 219 tgaccgcgaagaagcaaaagaaagctcttgacaagtagtgagctagaacacatgactaa 278

```

```

CNS0039G 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921 GI:4941778
VERSION AL063921
KEYWORDS fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pictet de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp. The same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source            1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT          201 a 64 c 131 g 202 t 503 others
ORIGIN
Query Match          5.6%; Score 65.4; DB 229; Length 1101;
Best Local Similarity 15.9%; Pred. No. 4.6e-05;
Matches 83; Conservative 249; Mismatches 187; Indels 2; Gaps 1;
Qy 601 ggtgagtgtaactctaggaaattatagaagatcttgcagcaacaatcccgcaatacaa 660
Db 1089 RDRTRKDDWTKWMTWKWKDRDRRWAGDADRWANDGAGTWTATWMTWMTWATWDTW 1030
Qy 661 aatatacgtttacgtcacgaacaagagacttaaaagcagatagagaatgcaatgaa 720
Db 1029 WDKWWWATAAKTDTATWMTWTRAWRADWAGRRDRGAGKRRDRDAATDADGAGRRDGRKKRD 970
Qy 721 gttgcaggaagagatatttaagagagcgtactctgttgatgacaggtta--cacaaattat 778
Db 969 KKDRKDDGDDKGGKKKAAKAAKWTWKWDDWDWDKDKWKGAKDRKADDDGGAGDKDD 910
Qy 779 atactaacaataatagtaataacataataatgacacaaagctggcagactgacctga 838
Db 903 DGKGRDADDDTGTGDKDDDDKDDWDDKAKGTWGTATWATAATDWWWGADADWMTWDA 850
Qy 839 gacaaagcgtgaatactcaaaagccttaaatgattggcgtgagagcgtgttacaagagt 898
Db 849 AADWWADDRWDWAWKWDADWAWGARTADRWDWGAGKRGKRGKRRDRKRRADDDKRAA 790
Qy 899 taaatatgctgcgcgactaggctcagacagatactctggaagaaagtcagagaacgtg 958

```



```

Query Match          5.6%; Score 64.8; DB 23; Length 853;
Best Local Similarity 78.0%; Pred. No. 6, le-05;
Matches 78; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

269 atgacttaaaactagaatgaagggttaaaactgagatgaagggttaaaactgaga 328
||||| |||||

```

T 14

Ph	Similarity	Score	64.8	DB	223	Length	899
78	Conservative	0	Mismatches	22	Indels	0	Gaps
gacgttaaaactaagaatgaagggttaaaactgagaatgaagggttaaaactgaga	328						
GAGT TAAACACAGAGATGAAGGTT TAAACACAGAGACGAAGGATTTAAACAGGAAA	442						

T 14

CNS00599/c	CNS00599	1036 bp	DNA	GSS	03-JUN-1999
LOCUS	Drosophila melanogaster	genome survey sequence	TET3 end of BAC #		
DEFINITION	BAC1116 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL057797				
VERSION	AL057797.1	GI:4932579			
KEYWORDS	GSS				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1036)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .				
FEATURES	Location/Qualifiers				
source	1..1036				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
	/clone_lib="RPCI-98"				
	/clone="BAC1116"				
	/note="end : TET3"				
BASE COUNT	64 a	56 c	41 g	701 t	174 others
ORIGIN					
Query Match	5.6%; Score 64.6; DB 229; Length 1036;				
Best Local Similarity	37.2%; Pred. No. 7e-05;				
Matches	262; Conservative 78; Mismatches 356; Indels 9; Gaps 2;				
QY	459	acaagctaataatgacaagtaccatggagagtgctgtatctatcattagcaccagaga	518		
DB	994	ACAAAGCRRMACGHCWCWYAAGCAWVTGTGARTATAAAATKGYGAWCKTRAWAGAK	935		
QY	519	taagctaaaaaattatgacgatcttgacgaacaaagacatgaattacaacagaa	578		
DB	934	TWAAAASAAASATCWAAGAGYGCASAAATWAAASAAATWARGWRAASACYGCAHAT	875		
QY	579	tgagaagttatctctgacgaagcgtgatggttaactcctagggaagttatagaagattctgc	638		
DB	874	SSAAAAGCWSACGWNKASACAYGAATAAAAAAAGAAKAGTCTCTGYCKCYGT	815		
QY	639	agcaaacaaatcccgcaatacaaaatagctttacgtcacgaacaaaggaact-taaaag	697		
DB	814	CTCAGKAKAGCYGYCAAAAAAAMATCCAAAGHAACAYAAAAAAGAAAGYGAAGTCTAAAT	755		
QY	698	cgaattagagaatgaatggaagttgcgaggaagattttaagagagcgtggtaccttgc	757		
DB	754	AGAAAATATYCCSWSAAHAKACTAGRAAAAAAAGAAAGAAVAAASGYHBAKTKK	695		
QY	758	tagatcaggttacacattatatacacaataatagtaattaccacaa-----ta	809		
DB	694	CKCAAMAMVTCNTAATAABKAKCTTWAAMAAAAAAMRCACAAAAAAGYGYSTS	635		
QY	810	taatgcacaagctggcagactgacctgagacaaaaggctgaatatctctaaaggccttaa	869		
DB	634	TYACATRAAAANRRRAAAARVWAAAAAATCGAAMAGAAAGAGTAA	575		
QY	870	tgattggctgagagctgttacagagagtaaatatcgatcgcgactagcttcagac	929		
DB	574	TAAYAGYAATGAKAWAATAWAKAAGTSTRAATATWAKKAAGAAANAKAGRTDGAKAA	515		
QY	930	agatactctggaaaagtacagaacgtgctgacaagtttgagatagataaacaatacgtt	989		
DB	514	AAAAAKAAATRTADAAAAAATAAAAKAKATGKBTBBAATAAAARAGADAWNKAGDG	455		
QY	990	aaacttaagaatagtgacttaagttttaataataaagcgttaaaagatacataatgatga	1049		
DB	454	GRGAAGKCKAGWATAAAATTAATGTATAAAAWGATKKKARAGAKAAAAAANAAN	395		
QY	1050	gttaactgaagcttagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagta	1109		
DB	394	ANNAAAAAAATAAAATAAAANNAANNAANNAANNAANNAANNAANNAANNAANNA	335		
QY	1110	tgaaaagctagtaataattcaagaattagaggcagcgttaagtaaaa	1154		
DB	334	AAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA	290		
RESULT	15				
CNS0182P					
LOCUS	CNS0182P	1101 bp	DNA	GSS	26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	ALI08811				
VERSION	ALI08811.1	GI:5629115			
KEYWORDS	GSS				
SOURCE	fruit fly.				
ORGANISM	Plasmid Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.				
FEATURES	Location/Qualifiers				
source	1..1101				
	/organism="Drosophila melanogaster"				
	/plasmid="pBelOBAC11"				
	/db_xref="taxon:7227"				
	/clone_lib="DrosBAC"				
	/clone="BACN37D10"				
	/note="end : SP6"				
BASE COUNT	274 a	268 c	128 g	73 t	358 others
ORIGIN					
Query Match	5.5%; Score 64.2; DB 229; Length 1101;				
Best Local Similarity	22.8%; Pred. No. 8.7e-05;				
Matches	117; Conservative 173; Mismatches 223; Indels 0; Gaps 0;				
QY	26	agacagatactctggagaaaagtacagaacgtgctgacaaqtttgagatagataaacaata	85		
DB	515	AGAGGAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA	574		

Search completed: May 13, 2001, 04:29:58
Job time: 9128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:07 ; Search time 21.02 seconds
(without alignments)
1041.557 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 1954

Sequence: 1 ACWVATRSQDTFLERVQERA.....KNDKSLSEKSIQIELEARK 383

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1954	100.0	383	20 Y04368	Streptococcal hexa
2	966	49.4	247	21 B03114	S. pyogenes hybrid
3	961	49.2	247	15 R50993	Recombinant M24-M5
4	932.5	47.7	274	15 R50994	Recombinant M24-M5
5	932.5	47.7	274	21 B03115	S. pyogenes hybrid
6	622	31.8	247	15 R50997	Recombinant M19-M6
7	622	31.8	247	21 B03119	S. pyogenes hybrid
8	622	31.8	343	15 R50999	Recombinant tetra
9	622	31.8	343	21 B03121	S. pyogenes hybrid
10	576	29.5	173	15 R50998	Recombinant M24-M5
11	576	29.5	173	21 B03120	S. pyogenes hybrid

12	392	20.1	287	15 R50992	Recombinant M24-M5
13	392	20.1	287	21 B03113	S. pyogenes hybrid
14	338	17.3	135	15 R51001	M24-M5-M6-M19-M3-M
15	337.5	17.3	135	21 B03124	S. pyogenes octava
16	335	17.1	305	15 R50996	Recombinant M24-M5
17	335	17.1	305	21 B03117	S. pyogenes hybrid
18	321	16.4	187	15 R50995	Recombinant M24-M5
19	321	16.4	187	21 B03116	S. pyogenes hybrid
20	267	13.7	67	15 R51000	M24-M5-M6-M19 mult
21	267	13.7	67	21 B03123	S. pyogenes hybrid
22	257.5	13.2	483	18 W08927	Type-6 M-protein.
23	256	13.1	441	14 R41780	Streptococcus pyog
24	249	12.7	441	10 P90955	M6 streptococcal p
25	220.5	11.3	327	18 W20901	Plasmodium falcipa
26	207.5	10.6	980	21 B18294	Restin protein seq
27	205	10.5	1392	20 Y06999	SSAL streptococcal
28	203	10.4	42	16 R74258	Human 160kD mediat
29	202	10.3	1427	12 R10534	Amino acid sequenc
30	198.5	10.2	2954	20 Y01632	Plasmodium falcipa
31	196	10.0	1979	21 B18171	Streptococcus uber
32	193.5	9.9	561	19 W63043	Sequence encoded b
33	190.5	9.7	753	21 B08316	Rattus norvegicus
34	189.5	9.7	100	6 P50295	Human mitotin. Ho
35	189	9.7	1886	19 W54241	Human cancer assoc
36	188.5	9.6	2482	16 R72826	Arabidopsis thalia
37	188.5	9.6	2482	19 Y23996	Arabidopsis thalia
38	188	9.6	472	21 Y23443	Arabidopsis thalia
39	187.5	9.6	979	21 G40045	Arabidopsis thalia
40	187.5	9.6	1019	21 G40044	Arabidopsis thalia
41	187.5	9.6	1058	21 G40043	Arabidopsis thalia
42	187.5	9.6	1342	21 G31251	Arabidopsis thalia
43	187.5	9.6	1382	21 G31250	Arabidopsis thalia
44	187.5	9.6	1421	21 G31249	Arabidopsis thalia
45	187.5	9.6	3248	17 R99795	Kinetochore protei

ALIGNMENTS

RESULT 1
Y04368
ID Y04368 standard; Protein; 383 AA.
XX
AC Y04368;
XX
DT 23-JUN-1999 (first entry)
XX
DE Streptococcal hexavalent M protein vaccine.
XX
KW Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A;
KW immunogenic; immune response; pathogen; fusion protein.
XX
OS Streptococcus sp.
XX
PN W09913084-A1.
XX
PD 18-MAR-1999.
XX
PF 14-SEP-1998; 98WO-US19100.
XX
PR 12-SEP-1997; 97US-0058635.
XX
(IDVA-) ID VACCINE.
XX
Dale JB;
XX
WPI; 1999-215066/18.
XX
N-PSDB; X33103.
XX
Immunogenic fusion protein derived from group A streptococci
XX
Example 1; Fig 7; 50pp; English.

ID R50993 standard; Protein; 247 AA.
 XX R50993;
 XX DT 02-NOV-1994 (first entry)
 XX DE Recombinant M24-M5-M6-M19.
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 XX Peptide 1..95
 FT /label= M24
 FT Peptide 96..155
 FT /label= M5
 FT Peptide 156..192
 FT /label= M6
 FT Peptide 193..247
 FT /label= M19
 XX
 PN WO9406421-A.
 XX
 XX 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 XX WPI; 1994-118122/14.
 DR N-PSDB; Q45215.
 XX
 XX New immunogenic hybrid proteins derives from streptococcal M proteins
 - induces opsonic antibodies, for protective immunisation against
 against multiple group A streptococci serotypes
 PS Disclosure; Fig 4; 67pp; English.
 XX
 CC The sequences given in R50992-1001 represent hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC PKK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX
 SQ Sequence 247 AA;

Query Match 49.2%; Score 961; DB 15; Length 247;
 Best Local Similarity 78.5%; Pred. No. 8.4e-58;
 Matches 194; Conservative 0; Mismatches 1; Indels 52; Gaps 1;

QY 3 MVATRSQDTTLEKVOERADKFEIENNTLKLKNSDLSPFNKALKDHNDELTEELSSNAKEKL 62
 DB 1 mvatrsqdttlekvgeradkfeienntliklksnslsfnnkalkdhnDELTEELSSNAKEKL 60
 QY 63 R-----CSAVTRG 70
 DB 61 rkndkslsekskqnelearkadlekalegammfstadsakikleakadlegsvtrg 120
 QY 71 TINDPQRAKEALDYELNHDLKTNEGKLTENEGKLTENEGKLTENEGKLTEDVDRVFP 130
 DB 121 tindpqrakealdkyelenhdltkknegkltknegkltknegkltknegkltedvdrvfpr 180
 QY 131 GTVENPDKARELLANKYDVNSMLQANNDKLPWRVRYTRHTPEDKLLKIIDDLDAKEHELQ 190
 DB 181 gtvenpdkarelllnkydvnsmlqanndnlpwrvrytrhtpedkllkiiddldakehelq 240
 QY 191 QONEKLS 197
 DB 241 qqnekls 247
 RESULT 4
 R50994
 ID R50994 standard; Protein; 274 AA.
 XX
 AC R50994;
 XX
 DT 02-NOV-1994 (first entry)
 XX
 DE Recombinant M24-M5-M6-M19 (linker variant).
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 XX Peptide 1..115
 FT /label= M24
 FT Peptide 116..124
 FT /label= Linker
 FT Peptide 125..184
 FT /label= M5
 FT Peptide 185..193
 FT /label= Linker
 FT Peptide 194..230
 FT /label= M6
 FT Peptide 231..239
 FT /label= Linker
 FT Peptide 240..274
 FT /label= M5
 XX
 PN WO9406421-A.
 XX
 XX 31-MAR-1994.
 PD
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 XX WPI; 1994-118122/14.
 DR N-PSDB; Q45216.
 XX
 PT New immunogenic hybrid proteins derives from streptococcal M proteins

PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes

PS Disclosure; Fig 6; 67pp; English.

XX The sequences given in R50992-1001 represent hybrid M proteins which
XX contain the M24-M5-M6 and/or M19 subunits. These multivalent
XX proteins were constructed using fragments of the 5' regions of emm
XX genes that were amplified by PCR, ligated in tandem and expressed in
XX pK223.3. The amplified regions pref. encode protective and not
XX tissue-cross-reactive epitopes, which can then be linked into one
XX protein molecule. The recombinant hybrid protein may contain 113
XX N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
XX 35 from M19. Each section is linked by 2 amino acids specified by
XX the respective restriction enzyme sites that were synthesised into
XX the primers used to specify the PCR product. Multivalent M proteins
XX such as this may be used for protective immunisation against group A
XX streptococci, which esp. cause rheumatic fever and rheumatic heart
XX disease. Humoral antibodies raised against these proteins do not
XX react with heart tissue antigens but are effective against many
XX different serotypes. The multivalent proteins may also include
XX sequences which induce mucosal antibodies and do not require coupling
XX to an immunogenic carrier.

SQ Sequence 274 AA;

Query Match 47.7%; Score 932.5; DB 15; Length 274;
Best Local Similarity 71.2%; Pred. No. 8e-56;
Matches 195; Conservative 0; Mismatches 0; Indels 79; Gaps 3;

Qy 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDLDELTEELSNAKEKL 62
Db 1 mvatrsqdtlekvqeradkfeienntklknsdlsfnnkalkdhndelteeslakekl 60
Qy 63 R-----
Db 61 rkndkslsekskngelkalekalegamnfstadsakiktleakadlegspgpnpa 120
Qy 64 --GSVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTEENGLKTENEGLK 121
Db 121 vpgsavgtrgtindpqrakealdkyelenhldlktkneglkteneglkteneglk 180
Qy 122 TE-----VDRVFPRTGTVENPKARELLNKYDVNSMLQANNDKL-----PWR 163
Db 181 tevdpnnpavpvdvrfvrgtvenpdkarellnkkydvnsmlqanndklpwpnnpavppwr 240
Qy 164 VRYTRHTPEDKLKKIIDDLDLDAKEHELOQONEKLS 197
Db 241 vrytrhtpedklkkliddldakehelqgqnekls 274

RESULT 5

B03115
ID B03115 standard; Protein; 274 AA.

XX AC B03115;

XX DT 10-OCT-2000 (first entry)

XX DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:6.

XX KW Multivalent hybrid M protein; group A streptococcus; serotype;
XX immunogenic; sero-specific antibody; streptococcal infection;
XX KW cross reactivity; vaccine; acute rheumatic fever; ARF;
XX KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
XX pneumonia.

XX OS Streptococcus pyogenes.

XX OS Synthetic.

XX PN US6063386-A.

XX

PD 16-MAY-2000.
XX 15-SEP-1997; 97US-0937271.
XX 16-SEP-1992; 92US-0945954.
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX Lederer JW, Dale JB;
XX WPI; 2000-364475/31.
XX N-PSDB; A57895.
PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX Disclosure; Fig 6A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
XX streptococcal M proteins comprising N-terminal peptide fragments of M
XX proteins that elicit opsonic antibodies against multiple serotypes
XX of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
XX generated using the hybrid proteins are against one or more M protein
XX serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
XX The invention also encompasses a recombinant DNA molecule comprising a
XX nucleotide sequence that encodes a multivalent hybrid M protein; and a
XX method for immunising a mammal against streptococcal infections,
XX comprising administering an immunogenic multivalent hybrid M protein to
XX the mammal. The multivalent hybrid M proteins are useful for eliciting
XX opsonic or protective antibodies to the M proteins of different serotypes
XX of group A streptococci, and may therefore be used as vaccines to protect
XX against and control infection by type A streptococci. Type A streptococci
XX are not only responsible for streptococcal pharyngitis (strep throat),
XX forms of pneumonia and a condition resembling toxic shock, but are also
XX involved in the development of acute rheumatic fever (ARF) and rheumatic
XX heart disease. In a patient with ARF, antibodies formed during a group A
XX streptococcal infection are also cross-reactive with heart tissue, which
XX indicates that the streptococci and host tissue contain similar antigenic
XX motifs. The new multivalent vaccines are capable of raising sero-specific
XX antibodies against various serotypes of group A streptococci which are
XX not cross-reactive with human heart tissue. Sequences B03113-B03117,
XX B03119-B03121 and B03123-B03124 represent multivalent hybrid
XX Streptococcus pyogenes M proteins generated in the disclosure of the
XX invention.

SQ Sequence 274 AA;

Query Match 47.7%; Score 932.5; DB 21; Length 274;
Best Local Similarity 71.2%; Pred. No. 8e-56;
Matches 195; Conservative 0; Mismatches 0; Indels 79; Gaps 3;

Qy 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDLDELTEELSNAKEKL 62
Db 1 mvatrsqdtlekvqeradkfeienntklknsdlsfnnkalkdhndelteeslakekl 60
Qy 63 R-----
Db 61 rkndkslsekskngelkalekalegamnfstadsakiktleakadlegspgpnpa 120
Qy 64 --GSVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTEENGLKTENEGLK 121
Db 121 vpgsavgtrgtindpqrakealdkyelenhldlktkneglkteneglkteneglk 180
Qy 122 TE-----VDRVFPRTGTVENPKARELLNKYDVNSMLQANNDKL-----PWR 163
Db 181 tevdpnnpavpvdvrfvrgtvenpdkarellnkkydvnsmlqanndklpwpnnpavppwr 240
Qy 164 VRYTRHTPEDKLKKIIDDLDLDAKEHELOQONEKLS 197
Db 241 vrytrhtpedklkkliddldakehelqgqnekls 274

Query Match	31.8%;	Score 622;	DB 15;	Length 247;
Best Local Similarity	39.2%;	Pred. No. 6.7e-35;		
Matches 149;	Conservative 10;	Mismatches 17;	Indels	204; Gaps
QY	4	VATRSQTDITLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLIR	63	
DB	40	vfrgtvnpdkarellinkdyvensmiganndnlv	74	
QY	64	GSAVTRGTINDPQRAKALDKYLENHDLKTKNEGLKTENEGLKTENEGKLTENEGKLTKE	121	
DB	75	-davtrgtindpgrakealdkyelenhdldktneglkteneglkteneglkteneglkte	131	
QY	124	VDRVFRPGTVENPDKARELLNKDYVENSMLQANDKLPWRVRYTRHTPDDEKKIIDDLD	181	
DB	134	pw	131	
QY	184	AKEHELOQONEKLSLQNGDGNPREVIEDLAANPAIGNIRLHENKDKLARENAMEYAG	241	
DB	136		131	
QY	244	RDFRRAGTLLDQVTLQYTKHNSYQOYNAQAGRLDLRQKAEYLKGLNDWAERLLQELNID	301	
DB	136		131	
QY	304	VATRSQTDITLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLIR	361	
DB	136	vatrsgtdtlekvqeradkfeientliklnsdlsfnnkalkdhndelteelsnakeklir	191	
QY	364	KNDKSLSERASKIOLEARK	383	
DB	196	kndkslsekskneleark	215	
RESULT	7			
ID	B03119	standard; Protein; 247 AA.		
AC	B03119;			
DT	10-OCT-2000	(first entry)		
DE	S. pyogenes	hybrid M protein (M19-M6-M5-M24), SEQ ID NO:13.		
KW	Multivalent	hybrid M protein; group A streptococcus; serotype;		
KW	immunogenic;	sero-specific antibody; streptococcal infection;		
KW	cross reactivity;	vaccine; acute rheumatic fever; ARF;		
KW	rheumatic heart disease;	streptococcal pharyngitis; strep throat;		
OS	Streptococcus	pyogenes.		
PN	US6063386-A.			
PD	16-MAY-2000.			
PF	15-SEP-1997;	97US-0937271.		
PR	16-SEP-1992;	92US-0945954.		
PA	(UYTE-)	UNIV TENNESSEE RES CORP.		
PI	Lederer	JW, Dale JB;		
DR	WPI;	2000-364475/31.		
PT	New immunogenic	recombinant hybrid M protein comprising amino-terminal		
PT	peptide fragments	of streptococcal M protein useful as vaccine against		
PT	rheumatic fever	and infections leading to rheumatic fever		
PS	Disclosure;	Fig 9A-B; 62pp; English.		
CC	The invention	relates to multivalent immunogenic hybrid group A		

[illegible]

QY 71 TINDPQRAKALDKYLENHDLTKNGLKTENEGKLTENEGKLTKE 123
 Db 121 tindpqrakealdkyelenhdltknegkltknegkltknegkltknegkltke 173

RESULT 12
 R50992
 ID R50992 standard; Protein; 287 AA.
 XX AC R50992;
 XX 02-NOV-1994 (first entry)
 XX DE Recombinant M24-M5-M6.
 XX KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Peptide 1..262
 FT /label= M24
 FT Peptide 263..275
 FT /label= M5
 FT Peptide 276..287
 FT /label= M6
 XX WO9406421-A.
 PN 31-MAR-1994.
 PD 15-SEP-1993; 93WO-US08703.
 XX 16-SEP-1992; 92US-0945954.
 PR (UYTE-) UNIV TENNESSEE RES CORP.
 PA Dale JB, Lederer JW;
 PI WPI; 1994-118122/14.
 DR N-PSDB; Q45214.
 XX New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX Disclosure; Fig 1; 67pp; English.

The sequences given in R50992-1001 represent hybrid M proteins which contain the M24-M5-M6 and/or M19 subunits. These multivalent proteins were constructed using fragments of the 5' regions of emm genes that were amplified by PCR, ligated in tandem and expressed in PKK223.3. The amplified regions pref. encode protective and not tissue-cross-reactive epitopes, which can then be linked into one protein molecule. The recombinant hybrid protein may contain 113 N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and 35 from M19. Each section is linked by 2 amino acids specified by the respective restriction enzyme sites that were synthesised into the primers used to specify the PCR product. Multivalent M proteins such as this may be used for protective immunisation against group A streptococci, which esp. cause rheumatic fever and rheumatic heart disease. Humoral antibodies raised against these proteins do not react with heart tissue antigens but are effective against many different serotypes. The multivalent proteins may also include sequences which induce mucosal antibodies and do not require coupling to an immunogenic carrier.

SQ Sequence 287 AA;
 Query Match 20.1%; Score 392; DB 15; Length 287;
 Best Local Similarity 100.0%; Pred. NO. 2.8e-19;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVOERADKFEIENNTLKNKNSDLSEFNKALKDHNDLTELSENAKEKLR 363
 Db 2 vatrsqtdtlekvgeradkfeienntlklksndlsfnkalkdhndelteeslnakeklr 61

QY 364 KNDKSLSEKASKIQLEFARK 383
 Db 62 kndkslsekaskiqeleark 81

RESULT 13
 B03113
 ID B03113 standard; Protein; 287 AA.
 XX AC B03113;
 XX 10-OCT-2000 (first entry)
 XX S. pyogenes hybrid M protein (M24-M5-M6), SEQ ID NO:2.
 DE Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.
 XX OS Streptococcus pyogenes.
 OS Synthetic.
 XX US6063386-A.
 PN 16-MAY-2000.
 PD 15-SEP-1997; 97US-0937271.
 XX 16-SEP-1992; 92US-0945954.
 PR (UYTE-) UNIV TENNESSEE RES CORP.
 PA Lederer JW, Dale JB;
 PI WPI; 2000-364475/31.
 DR N-PSDB; A57893.
 XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever
 XX Disclosure; Fig 1A-B; 62pp; English.

The invention relates to multivalent immunogenic hybrid group A streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19. The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a multivalent hybrid M protein; and a method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic

XX Streptococcus pyogenes.
OS Synthetic.
XX US6063386-A.
XX PD 16-MAY-2000.
XX 15-SEP-1997; 97US-0937271.
XX 16-SEP-1992; 92US-0945954.
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX Lederer JW, Dale JB;
XX WPI; 2000-364475/31.
XX N-PSDB; A57902.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX
XX Disclosure; Fig 13; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
CC invention.
XX
XX Sequence 135 AA;
SQ

Query Match 17.3%; Score 337.5; DB 21; Length 135;
Best Local Similarity 38.6%; Pred. No. 5.4e-16;
Matches 88; Conservative 6; Mismatches 21; Indels 113; Gaps 5;

QY 3 MVATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTELSNAKEKL 62
|||||
Db 1 mvatrsqtdtlekvqe----- 16

QY 63 RGSVTRGTINDPQRAKALDKYELENHDLTKNEGLKTENEGLKTENEGLKT 122
|||||
Db 17 -gsavtrgtindpgrak----- 32

QY 123 EYDRVFRPGTGVENPKARELLNKYDVENSMLQANNDKLPWRVYTRHPEDKLLKIITDDL 182
|||||
Db 33 evdrvfprgtvenpkar-----pwrvytrhtpedkllkl---q 69

QY 183 DAKEHELOQNEKLSLQDGNPREVIEDLAANNPAIQNIRLRHENKD 230
||| : ||||||| : | : |||

Db 70 darsvngfeprhvklihdngdgnpreviedlaaeafapl--trataadnkd 115

Search completed: May 10, 2001, 08:38:14
Job time: 67 sec

QY 244 RDKFRAGTLLDQVTLTKHNSNYQYNAQAGRLDLRQAEYKLGNDWAERLLQELNID 303
Db 136 ----- 135
QY 304 VATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 363
Db 136 ----- 135
QY 304 VATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 363
Db 136 ----- 135
QY 364 KNDKSLSEKASKIOLEARK 383
Db 196 KNDKSLSEKASKNOLEARK 215
RESULT 4
US-08-937-271-17
; Sequence 17, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-17
Query Match 31.8%; Score 622; DB 3; Length 343;
Best Local Similarity 39.2%; Pred. No. 1.7e-38;
Matches 149; Conservative 10; Mismatches 17; Indels 204; Gaps 3;
QY 4 VATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 63
Db 40 VPRGTVENPDKARELLNKYDVNSMLQANDNLV----- 74
QY 64 GSAVTRGTVNDPQRAKEALDKYLENHDLTKTNEGLKTENEGKLTENEGKLTKE 123
Db 75 -DAVTRGTVNDPQRAKEALDKYLENHDLTKTNEGLKTENEGKLTENEGKLTKE 133
QY 124 VDRVPRGTVENPDKARELLNKYDVNSMLQANDNLKLPWRVRYTRHPTEDKLLKIIDLD 183
Db 134 -----PW----- 135
QY 184 AKEHELOQNEKLSLQDGNPREVIEDLAANPAIONIRLRHENKDKIKARLENAMVAG 243

Db 136 ----- 135
QY 244 RDKFRAGTLLDQVTLTKHNSNYQYNAQAGRLDLRQAEYKLGNDWAERLLQELNID 303
Db 136 ----- 135
QY 304 VATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 363
Db 136 ----- 135
QY 364 KNDKSLSEKASKIOLEARK 383
Db 196 KNDKSLSEKASKNOLEARK 215
RESULT 5
US-08-937-271-15
; Sequence 15, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-15
Query Match 29.5%; Score 576; DB 3; Length 173;
Best Local Similarity 69.9%; Pred. No. 1.8e-35;
Matches 121; Conservative 0; Mismatches 0; Indels 52; Gaps 1;
QY 3 MWATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKL 62
Db 1 MWATRSQTDLTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKL 60
QY 63 R-----GSAVTRG 70
Db 61 RKNDKSLSEKASKNOLEARKADLEKALEGAMNFSTADSAKIKTLEAKADLEGSAVTRG 120
QY 71 TINDPQRAKEALDKYLENHDLTKTNEGLKTENEGKLTENEGKLTKE 123
Db 121 TINDPQRAKEALDKYLENHDLTKTNEGLKTENEGKLTENEGKLTKE 173

RESULT 6
US-08-937-271-2
; Sequence 2, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-2

Query Match 20.1%; Score 392; DB 3; Length 287;
Best Local Similarity 100.0%; Pred. No. 9.7e-22;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSENAKEKL 363
Db 2 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSENAKEKL 61
Qy 364 KNDKSLSEKASKIOLEARK 383
Db 62 KNDKSLSEKASKIOLEARK 81
RESULT 7
US-08-937-271-22
; Sequence 22, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-22
Query Match 17.3%; Score 337.5; DB 3; Length 135;
Best Local Similarity 38.6%; Pred. No. 3.8e-18;
Matches 88; Conservative 6; Mismatches 21; Indels 113; Gaps 5;
Qy 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSENAKEKL 62
Db 1 MVATRSQDTLEKVOE----- 16
Qy 63 RGSVTRGTINDPQRAKEALDKYELENHDLTKYNEGLKTENEGKTENEGKTENEG 122
Db 17 -GSVTRGTINDPQRAK----- 32
Qy 123 EVDVFPFGTGVENPKARELLNKKYDVENSMLQANDKLPWRVRYTRHTPEDKLLKIIDL 182
Db 33 EVDVFPFGTGVENPKAR-----PWVRYTRHTPEDKLLKL---Q 69
Qy 183 DAKEHLOQOONEKLSLQNGDGNPREVIEDLAANNPAIQNIRLHNK 230
Db 70 DARSVNGEPFRHVKLIDNGDGNPREVIEDLAAEFAPL--TRATADN 115
RESULT 8
US-08-937-271-10
; Sequence 10, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-475-6

Query Match 16.0%; Score 312; DB 2; Length 443;
Best Local Similarity 43.1%; Pred. No. 1.2e-15;
Matches 93; Conservative 18; Mismatches 65; Indels 40; Gaps 7;

QY 200 NCGNGPREVIEDLAANNPAIONIRLRHENKDKARLENAMVAGRDFKRA-----GTL 252
Db 1 NGDGNPREVIEDLAANNPAIONIRLRHENKDKARLENAMVAGRDFKRAEKEAKQAL 60
QY 253 LDQVTQLYTKHNSVQOY-----NAQAGRLDRLQKAEYLKGLNDWAER----- 295
Db 61 EDQRDKLETKLELQDDYDLAKESTSWDRQRLKELEBEKKEALSLAIDQASRDYHRATAL 120
QY 296 -----LLOELNIDVATR--SQTDTLEKVOERADK--FEIENNTL---KLKNSDLSFNN 341
Db 121 EKELEKKKALELAIDQASQDYNRANVLEKELETITREQEIENRNLGNNAKLELDQLSSEK 180
QY 342 KALKDHNDLTFEE--LSNAKEKLRKDKSLSEKASK 375
Db 181 EQLTTEKAKLEEKQISDASQSLRRDLSDASREAKK 216

RESULT 11
US-08-937-271-20
; Sequence 20, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 67 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-20

Query Match 13.7%; Score 267; DB 3; Length 67;
Best Local Similarity 38.3%; Pred. No. 2.3e-13;
Matches 67; Conservative 0; Mismatches 0; Indels 108; Gaps 3;

QY 3 MVATRSQTDITLEKVOERADKFEIENNTLKLKNSDLSFNNKALKDHNDLTELSELSNAKEKL 62
Db 1 MVATRSQTDITLEKVOE----- 16
QY 63 RGSAVTRGTINDPQRAKEALDKYLELHDLTKNEGLKTENEGKLKTENEGKLKT 122
Db 17 -GSAVTRGTINDPQRAK----- 32
QY 123 EVDVRFPRGTVENPDKARELLNKKYDVNSMLQANNDKLPWRVRYTRHTPTEDKLKK 177
Db 33 EVDVRFPRGTVENPDKAR-----EMRVRYTRHTPTEDKLKK 67
RESULT 12
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 9.6%; Score 188.5; DB 1; Length 2482;
Best Local Similarity 21.7%; Pred. No. 1.1e-05;
Matches 110; Conservative 92; Mismatches 145; Indels 159; Gaps 22;

QY 8 SQTDTLEKVOERADKFEIENNTLKLKNSDLSFNN--KALKD---HNDELTELSNAKEKL 62
Db 1424 NOIAQINKEKELLVK-ESESQARLSESDYEKLNVSKALEALVEKGFALRLSTQEEV 1482
QY 63 -----RGSAVTRGTINDPQRAK-----EALDKYLELHDLTKNEGLKTENEGKLKTENEG- 113
Db 1483 HOLRRGIEKLRVRIEADKKQLHIAEKLKERERENDSLKDKVENLERELQ-MSEENQELV 1541
QY 114 -----KTENEGLKTENVDRPFRGTVENPDKARELLNKKYDVNSMLQANNDKLPWRV-- 164
Db 1542 ILDAENSKAEVETLKTQIE-----EMARS-LKVFELDLVTLRSEKENLTQIQE 1589

TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 9.6%; Score 187.5; DB 5; Length 3248;
Best Local Similarity 21.7%; Pred. No. 1.8e-05;
Matches 110; Conservative 92; Mismatches 145; Indels 159; Gaps 22;

QY 8 SQTDTLEKVRADKFEFNTNKLKNSDLSFNN--KALKD---HNDELTELSNAEKL 62
DB 2152 NQIAQLNKEKELLVK--ESELQARLSSESDYEKLNVSKALEAALVEKGEFALRLSSTQEEV 2210
QY 63 ---RGSVTRGTINDPORAK---EALDKYLENHDLTKNEGLKTENEGLEL 113
DB 2211 HQLRGIEKLRVRIEADKKHIAELKEREENDSLKDKVENLERELQ--MSEENQELV 2269
QY 114 -----KTENEGLKTVEYRPPRGTVENPKARELLNKYDVNSMLQANNKLPWRV-- 164
DB 2270 ILDAENSKAEVETLKTQIE-----EMARS-LKIPELDVLTLRSEKENTKQIQE 2317
QY 165 RYTRHTPEDKL---KKIIDDLDKAE-----HELOQNEKLSLQONGGPN 205
DB 2318 KQGQSLDKLSSPKSLKEEQAEIQKEESKTAVEMLQNLQKELNEAVALCGD--- 2374
QY 206 RVIE--DLAANNPAIQLRHNENKDLKARLENAMVAGDFKFRAGTLLDQVTLTKH 263
DB 2375 QEIMKATEQSDDPPEEBHQLRNSTEKARLE-----ADERKKQCVLOQLKE--SEH 2425
QY 264 NSNY-----QQVNAQAGRLDLQKAEYKGLNDWAERLIQELNIDVAT 306
DB 2426 HADLLKGRVENLERELEARTNOEHAALAEANSKEVETLAKIEGTQSLRGLELDVVT 2485
QY 307 -RSQ---TDLEKVRADKFEFNTN---LKLKNSDL-----S 338
DB 2486 IRSEKENLTNELOKEQERISELEIINSFENILOKEQEQVMKEKSTAMEMLTQLKE 2545
QY 339 FNNKALKOHND-----LTEE 354
DB 2546 LNERVAALHNDQEAQKAEQNLSSQVECLEKALQLOGLDEAKNNYIVLOSSVKGLIQE 2605
QY 355 LSNAKEKLRKNDKSLSEKASKIQEEL 380
DB 2606 VEDGKQLEKDEISRLKNQIQDQE 2631

RESULT 15
US-08-630-822A-62
; Sequence 62, Application us/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDA
; TITLE OF INVENTION: NOVEL ECOTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C. Suite 3500
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630.822A
FILING DATE: 11-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-9223
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-822A-62

Query Match 9.5%; Score 185; DB 2; Length 900;
Best Local Similarity 24.2%; Pred. No. 5.8e-06;
Matches 104; Conservative 70; Mismatches 175; Indels 80; Gaps 19;

QY 7 RSQTDITLX-----VOERADKFEIENNTLKLKNSDLSFNNKALKDHNDLSELSNA 58
DB 423 RNQAGDYKELTKFKLSCKELQKAEFENELRRKTESLLVETKKRLDEQNKRTREMN 482
QY 59 KEKLRGSVTRGTINDPQRAKALDKYLENHDLTKNEGLK-----TE--NEGLKTE 109
DB 483 QOHNDKINLEKQINDLQ---EKL-KGELE-HNOKLKKQAVELRVAQSAATEQLNNELOET 537
QY 110 NEGLKTENEGLKTVEYRPPRGTVENPKARELLNKYDVNSMLQANNKLPWRVYTRH 169
DB 538 MQGLQTRDALQOEVASL--QGLKSQERSRSQASDMQIE---LEAKLOALHIELEHVN 592
QY 170 TPEDKLLKIIDDLDAKEHELOQNEKLSLQONGGPNPREVIEDLAANNPAIQLRHNENK 229
DB 593 C-EDKVTQ--DNQQLLERISTLEKECASLE-----LELKATQNKYEQEVKAHRETE 640
QY 230 DLKARLENAMVAGDFKFRAGTLLDQVTLQTYTKHNSNYQQVNAQAGRL----- 277
DB 641 --KSRLSVKSEANWEEVKALQIKLNEEKSAKROKSDQNSQEKERQISMLSVDYRQIQORLQ 698
QY 278 ----DLRQKAEYKGLNDWAEL-----RLQLNIDVATRSQTDTL-----EKVQ 317
DB 699 KLEGEYRQSESEKVKALHSQIEQEQKKSQIQSELGVQ---RSQTAHLTAREAQLVGEVAH 755
QY 318 ERADKFEIENNTLKLKNSDLSFNNKALKDHNDL--TEELSSNAKEKLRKND--KSLSEKAS 374
DB 756 LRDAKRNVEELHKLKTA-RSVDNAQMKLEQEQVEAEQVFSTLYKTHSNELKEEKS 814
QY 375 KIQELEARK 383
DB 815 HQEMEEER 823

Search completed: May 10, 2001, 08:37:51
Job time: 44 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:07 ; Search time 18.66 seconds
(without alignments)
1410.551 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 1954
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKAKIOLEARK 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 67: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397	20.3	539	2	A28549
2	385.5	19.7	492	2	A28616
3	339	17.3	197	1	MMSOMP
4	312	16.0	484	2	S46489
5	311	15.9	484	2	S35401
6	302	15.5	227	2	S01260
7	257.5	13.2	483	2	A26297
8	252	12.9	83	2	S61086
9	248	12.7	484	2	S34978
10	227	11.6	2139	2	T18296
11	220.5	11.3	327	2	F71980
12	214.5	11.0	150	2	S60838
13	212.5	10.9	1526	2	T41522
14	207.5	10.6	980	2	E71606
15	206.5	10.6	1931	2	A59234
16	206.5	10.6	2094	2	S33124
17	205	10.5	1392	2	A43336
18	204.5	10.5	520	2	S35575
19	204.5	10.5	911	2	S51441
20	204.5	10.5	1927	2	A59236
21	203	10.4	1827	2	T16270
22	203	10.4	1992	1	S02771
23	202	10.3	1427	2	S22695
24	200	10.2	1208	2	T39068
25	200	10.2	1939	1	A46762
26	199.5	10.2	1738	2	T14867
27	199	10.2	1509	1	A27224
28	198.5	10.2	2954	2	T14156
29	197	10.1	764	2	I51302

30 197 10.1 1496 2 T05634
31 197 10.1 1837 2 T41023
32 197 10.1 1939 2 T48175
33 196 10.0 1979 2 C71622
34 195.5 10.0 2288 2 T29999
35 195 10.0 1937 2 T38055
36 195 10.0 1938 2 I49464
37 193.5 9.9 1875 2 S38173
38 193.5 9.9 1934 2 T48153
39 193.5 9.9 1935 2 A59286
40 193.5 9.9 1957 2 T38077
41 193 9.9 875 2 S70115
42 193 9.9 1790 2 S67593
43 192.5 9.9 1935 1 A37102
44 192.5 9.9 1935 1 S06006
45 192 9.8 98 2 S61079

hypothetical prote
probable nuclear p
myosin heavy chain
hypothetical prote
hypothetical prote
myosin heavy chain
alpha cardiac myos
myosin-like protei
myosin heavy chain
hypothetical coile
ZIP1 protein - yea
transport protein
myosin beta heavy
myosin beta heavy
M protein precurs

ALIGNMENTS

RESULT 1
A28549
M24 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M24
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 26-Aug-1999
C:Accession: A28549; S60802
R:Mouw, A.R.; Beachey, E.H.; Burdett, V.
J. Bacteriol. 170, 676-684, 1988
A:Title: Molecular evolution of streptococcal M protein: cloning and nucleotide sequ
A:Reference number: A28549; MUID:88115166
A:Accession: A28549
A:Molecule type: DNA
A:Residues: 1-539 <MOD>
A:Cross-references: GB:M19031; NID:g153616; PIDN:AAA26874.1; PID:g153617
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60802
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 30-89 <WHA>
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein

Query Match 20.3%; Score 397; DB 2; Length 539;
Best Local Similarity 33.0%; Pred. No. 4.4e-12;
Matches 133; Conservative 57; Mismatches 155; Indels 58; Gaps 13;

QY 4 VATRSQDTLEKVOERADKFEIENNTLKNSDLSFNKKALKDHNDELTEELSNAKEKL 63
Db 43 VATRSQDTLEKVOERADKFEIENNTLKNSDLSFNKKALKDHNDELTEELSNAKEKL 102
QY 64 GSAVTRCTINPQRAKALDKYLE--NHDLKTKNEG-----LKTENEGUKTEN 110
Db 103 K-----NDKSLSEKAKIOLEARKADKALEGAMNFSTADSAKIKT---LEAEK 150
QY 111 EGLKTENEGUKTEVDVFPFGTVPENPKARELLNKYDVNSMLQANNDKLPWVRVTRH- 169
Db 151 AALAAKADKALEGAMNFSTA--DSAK--IKTLEAEKAALEAEKALEGAMNF 205
QY 170 --TPDKLKKIIDD---LDAKEHELOQONE-KUSLQNGDGNPREVIEDLAANNPAIONIR 223
Db 206 STADSAAKIKTLEAEKAALEAEKADKALEGAMNFSTADSAKIKTLE-----AEKAA 257
QY 224 LRHENDKLKALENAMVAGRDVFKRAGTLLDQVQTLT-----KHNSNYQQVNAQAGRLD 278
Db 258 LEARQAELEKALEGAMNFSTADSAKIKTLEAEKAALEAEKADKALEHQSQVNLAMROSRLRD 317
QY 279 LRQAEYLKGLNDWAERLLQELNIDVATR-SQTDLTLEKVOERADKFEIENNTLKNSDL 337

```
Db 318 LDASREAKKQLEAEHQLEQNKISEASRQSLRRDLDSAREAKKQLEAEHQLEEQ----- 373
Qy 338 SFNNKALKDHNDLETELSNAKEKLKRNKSLSEKASKIQELE 380
Db 374 ---NKISEASRQSLRRDLDSAREAKKQVERALEANSKLALE 413

RESULT 2
A28616
M5 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M5
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 26-Aug-1999
C:Accession: A28616; S60787
R:Miller, L.; Gray, L.; Beachey, E.; Kehoe, M.
J. Biol. Chem. 263, 5668-5673, 1988
A:Title: Antigenic variation among group A streptococcal M proteins. Nucleotide sequence
A:Reference number: A28616; MUID:88186881
A:Accession: A28616
A:Molecule type: DNA
A:Residues: 1-492 <MIL>
A:Cross-references: GB:M20374; NID:g153812; PIDN:AAA26976.1; PID:g153813
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the pop
A:Reference number: S60784; MUID:95198537
A:Accession: S60787
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: DNA
A:Residues: 30-89 <WHA>
C:Genetics:
A:Gene: smf5
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein
F:1-42/Domain: signal sequence #status predicted <SIG>
F:43-492/Product: M5 protein #status predicted <MAT>

Query Match 19.7%; Score 385.5; DB 2; Length 492;
Best Local Similarity 33.2%; Pred. No. 1.4e-11;
Matches 126; Conservative 47; Mismatches 106; Indels 101; Gaps 14;

Qy 52 TEELSNAKEKLGRSAVTRGTINDPORAKEALDKYLENHDLTKEGLKTENEGKLTENE 111
Db 37 TNEVS-----AAVTRGTINDPORAKEALDKYLENHDLTKEGLKTENEGKLTENE 88
Qy 112 GLKTENEGLKTEVDVPRGTVPNDPKARELLNKYDVNSMLQANNDKLPWRVRYTRHTP 171
Db 89 GLKTENEGLKTE-----KKEHE-----AENDKLK-QQDRTLTSTQ 121
Qy 172 EDKLLKIIDDLDAKEHELOQOQNEKLSLQNGD-----GNPREVIEDLAANNPA 218
Db 122 KETLEREVQN-----TQYNNETLKINGDLTKELNKTROELANKQOESKEKALNEL 174
Qy 219 IQ-----NIRLRHENKDLKARLENAMAVAGRD-----FKRAGTL---LDQVT----- 257
Db 175 LEKTVKDKTAKQENKETIGTLKKILDETVDKDKIAKEQENKETIGTLKKILDETVDKDKLA 234
Qy 258 -----QLYTHKNSNYQYNNQAG-RDLRLQKAPYKGLNDWAERLLQELNI 302
Db 235 KEQSKQNTGALKQELAKKDEANKISDSRKGRLRDLDSAREAKKQLEAEHQLEEQNKI 294
Qy 303 DVATRS-QTDTLEKVOERADKFEIENNTLKLKNSDLSFNKKALDKHDELTEELSNKAE 360
Db 295 SEASRGLRDR-LDASREAKKQLEAEHQKLEEQNKISEASRKGIR-----RDLDSARE 346
Qy 361 KLRKNDKLSLEKASKIQELE 380
Db 347 AKQVEKALEANSKLALE 366

RESULT 3
```

```
MMSOMP
M5 protein - Streptococcus pyogenes (fragment)
C:Species: Streptococcus pyogenes
C:Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Feb-1997
C:Accession: A03501
R:Manjula, B.N.; Acharya, A.S.; Mische, S.M.; Fairwell, T.; Fischetti, V.A.
J. Biol. Chem. 259, 3686-3693, 1984
A:Title: The complete amino acid sequence of a biologically active 197-residue fragment
A:Reference number: A03501; MUID:84162039
A:Accession: A03501
A:Molecule type: protein
A:Residues: 1-197 <MAN>
C:Comment: Residues 27-58 contain several tandem 7-residue repeats, and residues 150-
C:Comment: This protein is closely associated with the virulence of the bacterium and
C:Superfamily: M5 protein
C:Keywords: cell wall; transmembrane protein; virulence

Query Match 17.3%; Score 339; DB 1; Length 197;
Best Local Similarity 44.2%; Pred. No. 7.5e-10;
Matches 87; Conservative 23; Mismatches 49; Indels 38; Gaps 7;

Qy 67 VTRGTINDPORAKEALDKYLENHDLTKEGLKTENEGKLTENEGLKTEVDVDR 126
Db 2 VTRGTISDPORAKEALDKYLENHDLTKEGLKTENEGKLTENEGLKTENEGKLTENEGLKTE--- 58
Qy 127 VPRGTVPNDPKARELLNKYDVNSMLQANNDKLPWRVRYTRHTPEDKLLKIIDDLDAKE 186
Db 59 ---KSNLER--KTAELTS-----EKHEHEAENDKLK-QQDRTLTSTQKETLEREVQN----- 103
Qy 187 HELQOQNEKLSLQNGD-----GNPREVIEDLAANNPAIQ-----NIRLRHEN 228
Db 104 --TQYNNETLKINGDLTKELNKTROELANKQOESKEKALNELKTVRKDKIAKEQEN 161
Qy 229 KDLKARLENAMAVAGRD 245
Db 162 KETIGTLKKILDETVDK 178

RESULT 4
S46489
M1 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
C:Accession: S46489; S46490
R:Akesson, P.; Schmidt, K.H.; Cooney, J.; Bjoerck, L.
Biochem. J. 300, 877-886, 1994
A:Title: M1 protein and protein H: IgGc- and albumin-binding streptococcal surface p
A:Reference number: S46489; MUID:94280417
A:Accession: S46489
A:Molecule type: DNA
A:Residues: 1-484 <AKE>
A:Experimental source: strain 40/58, serotype M1
A:Accession: S46490
A:Molecule type: protein
A:Residues: 42-51 <AKW>
A:Experimental source: strain 40/58, serotype M1
C:Genetics:
A:Gene: emm1
C:Superfamily: M5 protein
C:Keywords: transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-484/Product: M1 protein #status experimental <MAT>
F:459-477/Domain: transmembrane #status predicted <TMM>

Query Match 16.0%; Score 312; DB 2; Length 484;
Best Local Similarity 43.1%; Pred. No. 3.9e-08;
Matches 93; Conservative 18; Mismatches 65; Indels 40; Gaps 7;

Qy 200 NGDGNPREVIEDLAANNPAIQNIRLRHENKDLKARLENAMAVAGRD-----GTL 252
Db 42 NGDGNPREVIEDLAANNPAIQNIRLRHENKDLKARLENAMAVAGRDPKRAEELEKAKQAL 101
```

Query Match	15.5%;	Score 302;	DB 2;	Length 227;	
Best Local Similarity	45.9%;	Pred. No. 4.9e-08;			
Matches	90;	Conservative	20;	Mismatches 58;	Indels 28; Gaps 9;
QY	200	NGDGNPREVIEDLAANPAQNTI	RLRHENKDLKARLENAM	EVAGDRPKRA-----GTL	252
Db	42	NGDGNPREVIEDLAANPAQNTI	RLRHENKDLKARLENAM	EVAGDRPKRAEELKAKQAL	101
QY	253	LDQVTOYLYTHKNSWYQYN--AQ	GRGLDRQAEYLGKLNDAW	ERLLQELNIDVATRS--Q	309
Db	102	EDQRKDLLETKELQPDYDLAK	ESTSDWRORLE--KELEEK	KEAL--ELAIQASRDVHR	157
QY	310	TDITLEVQERADKPEIENNTLK	NSDLSPFNKALKDHN--DE	LTELSNAKEKLKNDK	367
Db	158	ATALEK-----ELSEE--	KKALELAI--DQASQDYN	RANVLEKELETITRQE	205
QY	368	SLSEKASKIOELEARK	383		
Db	206	LLGNAKLELDQLSSEK	221		

RESULT	7
A26297	M6 protein - Streptococcus pyogenes
C:Species:	Streptococcus pyogenes
C:Date:	16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 26-Aug-1999
R:Accession:	A26297
R:Hollingshead,	S.K.; Fišchetti, V.A.; Scott, J.R.
J. Biol. Chem.	261, 1677-1686, 1986
A:title:	Complete nucleotide sequence of type 6 M protein of the group A streptococcus
A:Reference number:	A26297; MUID:86111835
A:Accession:	A26297
A:Molecule type:	DNA
A:Residues:	1-483 <MOL>
A:Cross-references:	GB:M11338; GB:M11415; NID:g153699; PIDN:AAA26920.1; PID:g153700
C:Genetics:	
A:Gene:	emm6
C:Superfamily:	M5 protein
C:Keywords:	coiled coil; transmembrane protein
Query Match	13.2%; Score 257.5; DB 2; Length 483;
Best Local Similarity	29.0%; Pred. No. 1.4e-05;
Matches 105; Conservative	40; Mismatches 134; Indels 83; Gaps 13;

```
QY 89 NHDLTNRKNEGLKTEENGLKTENEGKTEVDVPRGTVENPDKARELLNKYDV 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 HYSURKLLKGTASVAVALSVIGAGLVNTNEVSA--RVPRGTVENPDKARELLNKYDV 65
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 ENSMLQANDKLPWRYRTRHTPEDKLKTIIDDLDAKEHELOQOQNEKLSLQNGDGNPREV 208
||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 ENSMLQANDKLT-----TENNLTQDNKNLTENKNLTQDNKNLTEN--KNLTQD 115
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 IEDLAANNPAI--QNIIRLHNENKDLKARLENAMVAG--RDFKRA-GILLDQVDTQLYTK 262
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 KNLTITENKELKABENLTTENKGLTKKLSAEAEAEANKERENKEAIGTLTKLTIDETVKD 175
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 HNSNYQVYNAQAGRL-----DLRQKAEYLGKGLNDWAERLLQELNIDIVATRSOTD-- 311
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 KIAEQESKETIGTLTKLTIDETVKDKTAKQESKETIGTLTKLTIDETVKDKIAKQESKE 235
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 ---TLEKY-----QERADKFE---IENNTLKLKNSDLSFNKK 342
||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 TIGTLKILDTETVKDKIAREQSKQDIALGALKQELAKKDEGNKVSASRKGLRRDLASRE 295
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 ALKDHNDL---LTELNAEK-----LRKN-----DKSLSEKASKIQE 378
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 AKQVEKDLANLTAELDKVKEEKOISDASROGLRRDLASREAKKQVEKALEEANSKLAA 355
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 LE 380
||
Db 356 LE 357
||

RESULT 8
S61086
M protein precursor - Streptococcus pyogenes (serotype M68) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M68
C:Date: 15-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 16-Feb-1997
R:Whatmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
submitted to the EMBL Data Library, July 1994
A:Description: Noncongruent relationships between variation in emm1 gene sequences and the por
A:Reference number: S61072
A:Accession: S61086
A:Molecule type: DNA
A:Residues: 1-83 <WHW>
A:Cross-references: EMBL:U11997
R:Whatmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the por
A:Reference number: S60784; MUID:95198537
A:Accession: S60825
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 18-43,'A',45-78 <WHW>
A:Cross-references: EMBL:U11997
C:Genetics:
A:Gene: emm68
C:Superfamily: M5 protein
F:1-30/Domain: signal sequence (fragment) #status predicted <SIG>
F:31-83/Product: M protein (fragment) #status predicted <MAT>

Query Match 12.9%; Score 252; DB 2; Length 83;
Best Local Similarity 94.3%; Pred. No. 3.5e-06;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIQNIIRLHNENKDLKARLENAMVAGRDFKRA 252
||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 NGDGNPREVIEDLPANNPAIQNIIRLHNENKDLKARLENAMVAGRDFKRAEL 83
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
S34978
M1.1 protein precursor - Streptococcus pyogenes
```

```
C:Species: Streptococcus pyogenes
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S34978; S31966
R:Harbaugh, M.P.; Podbielski, A.; Huegl, S.; Cleary, P.P.
Mol. Microbiol. 8, 981-991, 1993
A:Title: Nucleotide substitutions and small-scale insertion produce size and antigen
A:Reference number: S34978; MUID:93360826
A:Accession: S34978
A:Molecule type: DNA
A:Residues: 1-484 <HAR>
A:Cross-references: EMBL:Z21845; NID:949401; PIDN:CAA79893.1; PID:949402
C:Genetics:
A:Gene: emm1.1
C:Superfamily: M5 protein
C:Keywords: membrane protein
F:1-42/Domain: signal sequence #status predicted <SIG>
F:43-484/Product: M1.1 protein #status predicted <MAT>

Query Match 12.7%; Score 248; DB 2; Length 484;
Best Local Similarity 37.0%; Pred. No. 4e-05;
Matches 80; Conservative 24; Mismatches 72; Indels 40; Gaps 7;

QY 200 NGDGNPREVIEDLAANNPAIQNIIRLHNENKDLKARLENAMVAGRDFKRA-----GTL 252
||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 NNDGRSRDVTETIAANTTVQNIIRLHNENKDLKARLENAMVAGRDFKRAEELKAKQAL 101
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 LDQVTQLYTKHNSYQY-----NAQAGRLDLROKAEYLGKGLNDWAER----- 295
||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 EDQRKDLTKLKELODYDLAKESTSDRQRLEKELEKEKALELAIDQASRDYHRATAL 161
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 -----LLQELNIDVATR--SQTDTLEKQVQERADK-FEINNTL----KLKNSDLSFNN 341
||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 EKELEKKKALELAIDQASQDYNNRANVLEKELEAITREQEIINRLLGNKRLDQLSSEK 221
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 KALKDHNDLTFE--LSNAKEKLRKNDKSLSEKASK 375
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 EQLTTEKALEEKEQISDASROSRLRDLASREAKK 257
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
T18296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T18296
R:Guillen, N.
submitted to the EMBL Data Library, February 1997
A:Reference number: Z18865
A:Accession: T18296
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-references: EMBL:L03534; NID:gl850912; PID:gl850913; PIDN:AAB48065.1
C:Genetics:
A:Gene: mhca
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 11.6%; Score 227; DB 2; Length 2139;
Best Local Similarity 23.2%; Pred. No. 0.0021;
Matches 117; Conservative 78; Mismatches 169; Indels 142; Gaps 21;

QY 8 SQTDTLEKQVQERADKFEINNTLKL-----KNSDLSFNNKALKDHNDLTFE 53
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1309 ADAELNKKKKAQSDK---ELNSLKAELEALTAKSVESKNKDSENEKAUSEEIDQANE 1365
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 54 ELSNAKEKLRGSAVTRGTINDPQ-----RAKEALDKYLEEN 89
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1366 KLNQIADLRKATADLQEAANEKKAVEAQRDKLVADNKKMTKTLEIKARDEENTYKVEN 1425
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 HD--LTKNEGLKTEENGLKTEENGLK-----TEVDRVFP- 129
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 1426 YEKVLKRKEADLEAEANENLDIEKKDRMKKQVKKLEGELKETKDKLNAIAEKDSIFTA 1485
Qy 130 -----RGTVENPKARELLNKYDVENSMLQANNKLPWRVRYTHPTEDKLKKI 178
Db 1486 KQSDADLEELNKTVESHD---EVVALNTQITKLTRDNQSAEEELNELR-SKADKDKK 1541
Qy 179 IDDLDAKEHELOQOONKLSLQNGDGNPREV-----IEDL--AANNPAIQNIRLRHENKDLK 232
Db 1542 ISELEEQVNEL--ESRPVGTGNADENEIKIRDAQIADLNKALEMKGVQNNQLQATNKELK 1599
Qy 233 A-----RLENAM-----VAGRDFKRACTLLDQVTLQYTKH 263
Db 1600 AKNDLTSKIBITENEMKLENKAKRLEQDKDEADKAVSQTIKRG-LEEEVKKLTTEI 1658
Qy 264 NSNYQOYNAQAGRLDLRQAEYLKGLNDWAERLLLOELNIDVATRSQTDITLKVQERAD-- 321
Db 1659 QALFKQINAPSSVAQEERKORLESIDAELEKEQLEQERT--TAANAER-KKIQAELEDEV 1715
Qy 322 KFEIENNT-----LKNKSD-----LSPNNKALKDH-----NDELTEELSNKAK 361
Db 1716 KFNLEDVTNQREKLVAKNSENDAEIDSLEKEKALEDEIEKITDDNNKLESEEDSLDRKY 1775
Qy 362 ---LRKNDKSLSEKASKIQLEARK 383
Db 1776 NALLDSKSDSVSMKEKQDELKVTK 1800
RESULT 11
F71980
hypothetical protein jhp0052 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: F71980
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <ARN>
A:Cross-references: GB:AE001444; GB:AE001439; NID:g4154549; PIDN:AAD05629.1; PID:g415455
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0052

Query Match 11.3%; Score 220.5; DB 2; Length 327;
Best Local Similarity 25.7%; Pred. No. 0.00051;
Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;
Qy 14 EKVOERADKFEIENNTLKNKSLSPNNKALKDHNDL-----TEELSNKAKRGRSAVTRG 70
Db 16 EELARISELEDENTELLREYLAETSELKDANDQLRQKNDKLFITDKL----- 67
Qy 71 TINDPQRAKALDYELNHDLYTKNEGLKTENEGLKTENEGLKTENEGLKTENEGLKTENEGLKT 130
Db 68 -----TKNETLFA-ENESLSYKISGLSHSNDQLWQNNKLTKEKALTE----- 112
Qy 131 GTVENPKARELLNKYDVENSMLQANNKLPWRVRYTHPTEDKLLKIIDDLDAKEHELO 190
Db 113 -----KDILAK-----ENRLLAARDRLTEKRELT-TEKERLARENTELTHTKLTET 159
Qy 191 QONKLSLQNGDGNPREVIEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRAK 250
Db 160 KENKALATEN-----DKLNHQVTALTNERDSLEQ-----RARLQDA-----HG 198
Qy 251 TLLDQVTLQYTKHNSYQOYNAQAGRLDLRQAEYLKGLNDWAERLLLOELNIDVATRSQT 310
Db 199 FLEKRCNTNL-EKEN---ORUTDKLKQLESQAOKS--LENTNNQLRQALENSVLOAQAKEK 252

Qy 311 DTLEKVOERADKFEIENNTLKNK-----SDLSFNKALKDHNDLTEELSNKAK-- 361
Db 253 IAIE-----KSELERETARLUKSLGLEMAKSDLDLHNRRLASANEKLRQNRKLEENI 305
Qy 362 -LRKNDKSLSEKASKIQ 377
Db 306 ALKERVVDGLNEQLSKIQ 322
RESULT 12
S60838
M protein precursor - Streptococcus pyogenes (serotype M6) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M6
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60838
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-150 <WHA>
A:Cross-references: EMBL:U11986; NID:g533649; PIDN:AAA99602.1; PID:g1235835
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Superfamily: M5 protein

Query Match 11.0%; Score 214.5; DB 2; Length 150;
Best Local Similarity 43.6%; Pred. No. 0.0004;
Matches 61; Conservative 11; Mismatches 25; Indels 43; Gaps 6;
Qy 126 RVFPRTGVENPKARELLNKYDVENSMLQANNKLPWRVRYTHPTEDKLLKIIDDLDAK 185
Db 30 RVFPRTGVENPKARELLNKYDVENSMLQANNKLT-----TTENK----- 69
Qy 186 EHELOQOONKLSLQNGDGNPREVIEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRD 245
Db 70 --NLTDQKNLTDQN-----KELKA-----EENRLTTENKGLTKKLSAEAEAAANK 113
Qy 246 FRRA-----GTL---LDQVQT 258
Db 114 EQESKETIGTLAKTLDEIVK 133

RESULT 13
T41522
myosin ii - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
A:Accession: T41522
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z22000
A:Accession: T41522
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1526 <WOO>
A:Cross-references: EMBL:AL049498; PIDN:CAB39901.1; GSPDB:GN00068; SPDB:SPCC645.05c
A:Experimental source: strain 972h-; cosmid c645
C:Genetics:
A:Gene: SPDB:SPCC645.05c
A:Map position: 3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:80-743/Domain: myosin motor domain homology <MWO>

Query Match 10.9%; Score 212.5; DB 2; Length 1526;
Best Local Similarity 22.1%; Pred. No. 0.007;
Matches 93; Conservative 74; Mismatches 161; Indels 93; Gaps 12;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:52 ; Search time 11.75 Seconds
(without alignments)
1116.583 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 1954

Sequence: 1 ACMVATRSQDTFLKQVQERA.....KNDKSLSEKASKIQLEARK 383

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397	20.3	539	1 M24_STRPY	P12379 streptococ
2	385.5	19.7	492	1 M5_STRPY	P02977 streptococ
3	237.5	13.2	483	1 M6_STRPY	P08089 streptococ
4	212.5	10.9	1433	1 REST_CHICK	O42184 gallus gall
5	212.5	10.9	1526	1 MYS2_SCHPO	Q9usi6 schizosacch
6	206.5	10.6	2349	1 TPR_HUMAN	P12270 homo sapien
7	203	10.4	1969	1 MYS4_CAEEL	P12844 caenorhabdi
8	202	10.3	1427	1 REST_HUMAN	P30822 homo sapien
9	201.5	10.3	1102	1 MYSC_CHICK	P29616 gallus gall
10	200	10.2	1208	1 YDH6_SCHPO	Q92351 schizosacch
11	200	10.2	1939	1 MYS4_HUMAN	P13333 homo sapien
12	199	10.2	1509	1 MYSN_ACACA	P05659 acanthamoeb
13	197	10.1	1939	1 MYS4_MESAU	P13339 mesocricetu
14	195	10.0	1937	1 MYSP_HUMAN	P13355 homo sapien
15	195	10.0	1938	1 MYS4_MOUSE	Q02566 mus musculu
16	193.5	9.9	1875	1 MLPI_YEAST	Q02455 saccharomyc
17	193.5	9.9	1934	1 MYSB_MESAU	P13540 mesocricetu
18	193.5	9.9	1935	1 MYSB_PIG	P79293 sus scrofa
19	193.5	9.9	1957	1 YD86_SCHPO	Q10411 schizosacch
20	193	9.9	875	1 ZIP1_YEAST	P31111 saccharomyc
21	193	9.9	1790	1 USOI_YEAST	P25386 saccharomyc
22	192.5	9.9	1935	1 MYS4_HUMAN	P12883 homo sapien
23	192.5	9.9	1935	1 MYSB_RAT	P02564 rattus norv
24	191.5	9.8	2245	1 MYSJ_DICDI	P54697 dictyosteli
25	191.5	9.8	2663	1 CENE_HUMAN	Q02224 homo sapien
26	190	9.7	944	1 NUFL_YEAST	P32380 saccharomyc
27	189.5	9.7	866	1 MYSF_SCHJA	Q05870 schistosoma
28	189.5	9.7	2022	1 ANTI_ORCVO	P21249 onchocerca
29	189	9.7	866	1 MYSF_SCHMA	P06198 schistosoma
30	189	9.7	1928	1 MYS1_YEAST	P08964 saccharomyc
31	189	9.7	1938	1 MYS4_RAT	P02563 rattus norv
32	188.5	9.6	3210	1 CENE_HUMAN	P49454 homo sapien
33	187.5	9.6	2649	1 BPAL_HUMAN	Q03001 homo sapien

34	186.5	9.5	756	1 Y328_MYCGE	Q49419 mycoplasma
35	186.5	9.5	1972	1 MYST_RABIT	P35748 oryctolaqus
36	186.5	9.5	1976	1 MYSO_HUMAN	P35580 homo sapien
37	186	9.5	465	1 MYS4_RABIT	P04460 oryctolaqus
38	186	9.5	2116	1 MYS2_DICDI	P08799 dictyosteli
39	185	9.5	976	1 SCPI_HUMAN	Q15431 homo sapien
40	184.5	9.4	1033	1 Y328_MYCPN	P75310 mycoplasma
41	184	9.4	539	1 MYS3_HYDAT	P39922 hydra atten
42	184	9.4	863	1 MYSP_TAESO	P35418 taenia soli
43	184	9.4	1940	1 MYSE_CHICK	P03565 gallus gall
44	183.5	9.4	692	1 MYS_PODCA	Q05D00 podocoryne
45	183.5	9.4	1539	1 Y373_HUMAN	O15078 homo sapien

ALIGNMENTS

RESULT 1					
M24_STRPY	M24_STRPY	STANDARD;	PRT;	539 AA.	
ID	P12379;				
DT	01-OCT-1989 (Rel. 12, Created)				
DT	01-OCT-1989 (Rel. 12, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	M PROTEIN, SEROTYPE 24 PRECURSOR.				
GN	EM24.				
OS	Streptococcus pyogenes.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1314;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-A24 VAUGHN;				
RX	MEDLINE=88115166; PubMed=3276665;				
RA	Mouw A.R., Beachey E.H., Burdett V.;				
RT	"Molecular evolution of streptococcal M protein: cloning and				
RT	nucleotide sequence of the type 24 M protein gene and relation to				
RT	other genes of Streptococcus pyogenes.";				
RL	J. Bacteriol. 170:676-684(1988).				
CC	!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES				
CC	OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF				
CC	THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO				
CC	PHAGOCYTOSIS.				
CC	!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.				
CC	!- SIMILARITY: TO OTHER M PROTEINS.				
CC	!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS				
CC	IN THE REGION OF THE MEMBRANE ANCHOR.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/ or				
CC	send an email to license@isb-sib.ch)				
CC	-----				
EMBL:	M19031; AA226874.1; -				
PIR:	A28549; A28549.				
InterPro:	IPR001899; -				
Pfam:	PF00746; Gram_pos_anchor; 1.				
PRINTS:	PR00015; GPOSANCHOR.				
PROSITE:	PS00343; GRAM_POS_ANCHORING; 1.				
Signal:	Virulence; Cell wall; Transmembrane; Phagocytosis;				
Duplication;	Repeat; Antigen; Coiled coil.				
SIGNAL	1 42				
FT CHAIN	43 539				
FT DOMAIN	43 514				
FT TRANSMEM	515 534				
FT DOMAIN	535 539				
FT DOMAIN	118 301				
FT REPEAT	118 152				
FT REPEAT	153 187				
FT REPEAT	188 222				

```
FT REPEAT 223 257 A-4 (1 ALTERATION).
FT REPEAT 258 292 A-5 (2 ALTERATIONS).
FT REPEAT 293 301 A-6 (INCOMPLETE).
FT DOMAIN 311 405 2.7 X 35 AA TANDEM REPEATS, B-TYPE.
FT REPEAT 311 355 B-1.
FT REPEAT 356 380 B-2.
FT REPEAT 381 405 B-3 (INCOMPLETE).
FT DOMAIN 468 504 GLY/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 505 510 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
                          PROTEINS.
SQ SEQUENCE 539 AA: 58804 MW: 803EDF3AC1B6E9C7 CRC64;

Query Match      20.3%; Score 397; DB 1; Length 539;
Best Local Similarity 33.0%; Pred. No. 4.5e-12;
Matches 133; Conservative 57; Mismatches 155; Indels 58; Gaps 13;

Qy 4 VATRSDTLEKVGQERADKFEIENNTLKNLSLSPFNKALKDHNDELTELSNAKEKLR 63
Db 43 VATRSDTLEKVGQERADKFEIENNTLKNLSLSPFNKALKDHNDELTELSNAKEKLR 102
Qy 64 GSATRTGNDPQAKALDKYLE--NHDLYTKNEG-----LKTENEGKLTEN 110
Db 103 K-----NDKLSKASKIQEAEKADLEKALEGAMNFSTADSAKIKT-----LEAEK 150
Qy 111 EGLKTENEGKLTEDVFPFGTVPENPKARELLNKYDVNSMLQANNDKLPWRVRYTRH- 169
Db 151 AALAAKADLEKALEGAMNFSTA---DSAK--IKTLEAKAALAEKALEKALEGAMNF 205
Qy 170 --TPEDKLKKIIDD---LDAKEHELOQONE-KLSLQNGDGNPREVIDEALANPAIQNR 223
Db 206 STADSAKIKTLEAEKAAALAAKADLEKALEGAMNFSTADSAKIKTLE-----AEKAA 257
Qy 224 LRHENKDLKARLENAMVAGDFRAGTLLDQVTLT-----KHSNYQOYNAQGRILD 278
Db 258 LEARQAELEKALEGAMNFSTADSAKIKTLEAEKAALEAEKADLEHQSVLNARQSLRRD 317
Qy 279 LRQAEYLLKGNDAERLLQELNIDVATR-SQTDLEKVGQERADKFEIENNTLKNLSL 337
Db 318 LDASREAKKQLEAEHQLEEQNKISEASRQSLRDLDSAREAKKQLEAEHQLEEQ---- 373
Qy 338 SFNNKALKDHNDLTELSENAKEKLRNKSLSKASKIQELE 380
Db 374 ---NKISEASRQSLRDLDSAREAKKQVEKALEEANSKLALE 413
```

RESULT 2

```
ID M5_STRPY STANDARD; PRT; 492 AA.
AC P02977;
DC 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE M PROTEIN, SEROTYPE 5 PRECURSOR.
GN EM5 OR SMP5.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88186881; PubMed=3281944;
RA Miller L., Gray L., Beachey E., Kehoe M.;
RT "Antigenic variation among group A streptococcal M proteins.
RT Nucleotide sequence of the serotype 5 M protein gene and its
RT relationship with genes encoding types 6 and 24 M proteins.";
RL J. Biol. Chem. 263:5668-5673(1988).
RN [2]
RP SEQUENCE OF 43-212 AND 238-250.
RX MEDLINE=84162039; PubMed=6368549;
RA Marjula B.N., Acharya A.S., Mische S.M., Fairwell T., Fischetti V.A.;
RT "The complete amino acid sequence of a biologically active
RT 197-residue fragment of M protein isolated from type 5 group A
```

```
RT streptococci.";
RL J. Biol. Chem. 259:3686-3693(1984).
CC !- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC !- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC !- SIMILARITY: TO OTHER M PROTEINS.
CC !- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M20374; AAA26976.1; -.
DR PIR; A03501; MWSOMP.
DR PIR; A28616; A28616.
DR InterPro; IPR001899; -.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
KW Transmembrane; Coiled coil; Signal.
FT SIGNAL 1 42
FT CHAIN 43 492 M PROTEIN, SEROTYPE 5.
FT DOMAIN 43 466 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 467 486 MEMBRANE ANCHOR.
FT DOMAIN 487 492 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 421 457 GLY/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 458 463 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
                          PROTEINS.
FT DOMAIN 69 103 5 X 7 AA TANDEM REPEATS OF L-K-T-K-N-E-G.
FT REPEAT 69 75 1.
FT REPEAT 76 82 2.
FT REPEAT 83 89 3.
FT REPEAT 90 96 4.
FT REPEAT 97 103 5.
FT CONFLICT 43 43 A -> T (IN REF. 2).
FT CONFLICT 50 50 N -> S (IN REF. 2).
FT CONFLICT 102 102 K -> SNLERKTAELTSEK (IN REF. 2).
FT CONFLICT 208 208 I -> L (IN REF. 2).
SQ SEQUENCE 492 AA: 55085 MW: 873779B6CBD55E27 CRC64;
```

```
Query Match      19.7%; Score 385.5; DB 1; Length 492;
Best Local Similarity 33.2%; Pred. No. 1.4e-11;
Matches 126; Conservative 47; Mismatches 106; Indels 101; Gaps 14;

Qy 52 TEELSNKAEKLRGSATRTGNDPQAKALDKYLENHDLYTKNEGKLTENEGKLTENE 111
Db 37 TREVS-----AAVTRGNDPQAKALDKYLENHDLYTKNEGKLTENEGKLTENE 88
Qy 112 GLKTENEGKLTEDVFPFGTVPENPKARELLNKYDVNSMLQANNDKLPWRVRYTRHP 171
Db 89 GLKTENEGKLT-----KKEH-----AENDKLLK-QQDFTLSTQ 121
Qy 172 EDKLLKIIDDLEAEKHELOQONEKLSLQNGD-----GNPREVIDEALANPA 218
Db 122 KETLEREVQN-----TOYNNETLKKIKNGDLYTKNEGKLTENEGKLTENE 174
Qy 219 IQ-----NRLRHNKDLKARLENAMVAGRD-----FRAGTLL---LDQVT----- 257
Db 175 LEKTVKDKIAKEQENKETIGTIGLKILDETVDKDKIAKEQENKETIGTIGLKILDETVDKDKLA 234
Qy 258 -----OLYTKHSNYSQYNAQAG-RLDLRQKAEYLLKGNDAERLLQELN 302
Db 235 KEQKSKQNTGALKQELAKDEANKISDASRGLRDLDSAREAKKQLEAEHQLEEQNKI 294
```


CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF014012; AAC60344.1; -
DR EMBL; AF020764; AAC60345.1; -
DR EMBL; AF045650; AAC03547.1; -
DR EMBL; AF045651; AAC03548.1; -
DR InterPro: IPR000938; -
DR Pfam: PF01302; CAP_GLY; 2;
DR PROSITE; PS00845; CAP_GLY_1; 2;
KW Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
FT DOMAIN 79 121
FT SER-RICH. 144 207
FT CAP-GLY. 235 277
FT DOMAIN 305 332
FT SER-RICH. 351 1353
FT DOMAIN 1414 1427
FT CCHC-BOX.
FT VARSPPLIC 458 492
FT VARSPPLIC 458 492
FT MISSING (IN SHORT ISOFORM).
FT TQTKLHARIKELEQSLFEKTKADKLORELEDR ->
FT RKQISEDPEN (IN ISOFORM CLIP-170(11)).
FT S -> GGSSKVS (IN ISOFORM CLIP-170(11)).
FT T -> RKRQISEDPEN (IN ISOFORM CLIP-
FT 170(11+35)).
FT CONFLICT 309 309 K -> R (IN AAC03547).
FT CONFLICT 440 440 E -> V (IN AAC03548).
SQ SEQUENCE 1433 AA; 161026 MW; 5631CB8683498E23 CRC64;

Query Match 10.9%; Score 212.5; DB 1; Length 1433;
Best Local Similarity 21.4%; Pred. No. 0.0041;
Matches 92; Conservative 94; Mismatches 181; Indels 63; Gaps 13;

QY 5 ATRSDTDTLEKVRADKFEIENNTLAKKSDLSFNKALKDHNDLTELSENAKELRG 64
DB 854 ARAWQETINKNGEQFALMSSELSQLSNLTVMETKLE-REEREQQLTEAKVKLEN 912
QY 65 -----SAVTRGTINDPORAKE-ALDKYELENHDLTKNEGL-----KTENE 104
DB 913 DIAEIMKSSGSSAQLMKMNDLRLKERQLEIQLE---LTKANERAVQLQKNVEQTAQK 969
QY 105 GLKTENEGKLTENGLTEVDV--FPRGVENPDKARELLNKYDVENSMLQANDKLPW 162
DB 970 AEQSQQETLTHQBELKMDQDLTDMKKOMETSONQYKDLQAKYKETSEMITKHDADIK 1029
QY 163 RVRYTRHPEDKLLK---KIIDLDLDAKHEHQEQNEKLSLQNGDGNPREVIEDL----- 212
DB 1030 GFKNLLDAEALKAQKKDELETAQELKQAEQ---AKADKRAEEVLQTMKVTKEK 1086
QY 213 -AANPAIQIRLHENKDLKARLENAMVAGDFKRAGLTLLDQVTLTKHNSNYQQYN 271
DB 1087 DAHQEIETLASLENSQTNKELQNELMDLQNNLKNNEELTKSKELNLNENKKNVEELK 1146
QY 272 A--QAGRLDRQAEYKLGUNDAERLLQELNTDVAIRSDTDTLEKVRADKFEIENNT 329
DB 1147 KEFEALKAQAQSQQAALQAEENKVLAEELG---RSRDEVTSHQLEE--ERSVLNNQL 1201
QY 330 LKLNKSLSF-----NNKALKDHNDLTELSENAKELRKNDKSLSEKA 373
DB 1202 LEMKKRESTLUKKEIDERASLQKSI SDTSALITQKDEELEKLRNETVTLRGENASATLQ 1261
QY 374 SKIQELEARK 383
DB 1262 SVKLTLESDK 1271

RESULT 5
MYS2_SCHPO STANDARD; PRT; 1526 AA.
ID Q9US16; P78969;
AC DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN II HEAVY CHAIN.
GN MYO2 OR SPCC645.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=98075862; PubMed=9415380;
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "Type II myosin involved in cytokinesis in the fission yeast,
RL Schizosaccharomyces pombe";
RN Cell Motil. Cytoskeleton 38:385-396(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CC CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75357; AAC49908.1; -
DR HSSL; AL049498; CAB39901.1; -
DR HSPP; P08799; 1MND.
DR InterPro; IPR000048; -
DR InterPro; IPR001609; -
DR PRINTS; PR00193; MYOSINHEAVY.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; ATP-binding; Alkylation.
FT DOMAIN 1 815
FT GLOBULAR HEAD (S1).
FT DOMAIN 816 1526
FT RODLIKE TAIL.
FT DOMAIN 758 787
FT IQ.
FT DOMAIN 875 1244
FT COILED COIL (POTENTIAL).
FT NP_BIND 170 177
FT ATP (POTENTIAL).
FT DOMAIN 634 656
FT ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 734 748
FT ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 674 674
FT ALKYLATION (BY SIMILARITY).
FT CONFLICT 1337 1337 S -> R (IN REF. 1).
SQ SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;

Query Match 10.9%; Score 212.5; DB 1; Length 1526;
Best Local Similarity 22.1%; Pred. No. 0.0044;
Matches 93; Conservative 74; Mismatches 161; Indels 93; Gaps 12;

QY 11 DTLEKVRADKFEIENNTLKNKSDLSFNKALKDHNDLTELSENAKELRGSAVTRG 70
DB 880 ETLRRTQERLANIEDSFSETKQONLQRESASLKQINNESEL-----LEKTSKVT 933
QY 71 TINDPQRAEKALDKYELENHDLTKNEGLKTENEGKTE---NEGLKTENEGKLTVD 126
DB 934 LLSEQNELKELSLSEKLDLDTKGELESRENNATVLSKAEFNEQCKSLQETIVTKDAE 993

QY 127 VPRGTENPKARELLNKKYDVNSMLQANNDKLPWRVYRTRTPEDKLLKI-----178
 Db 994 L-----DKLKYSIDYKTEQEMRLNOKMNEKSIOQEGSLSRVKKLERENS 1044
 QY 179 --IDDLDAKEHELOOQNEKLSQNGDNPNREVIDLAANPAIONIRLHENKDKLAKLE 236
 Db 1045 TLISDVST-----LKOQKEELSVLKG-----VOELTINN-----LE 1075
 QY 237 NAMEVAGRDKFRAGTCL-----LDQVTOLYTKHNSNYQQ-----YNAQAGRDLRLQ 281
 Db 1076 EKVNYLEADYKQLPKLKELESNDKDLQLOQATKNKELEAKVKECLNNIKSLTKELEN 1135
 QY 282 KAEYLGNDWAERLLQELNIDVATRSQDTLEKQVPRADKFEIENNTLK-----331
 Db 1136 KEERQCNLSASLKYIELQETHENLLKVSDELENKKYEGQLDLEGLKDVDTNFOELS 1195
 QY 332 LKNSDLSFNKAL-----KHNDDEL-----EELSNAKEKLRKNDKLSKSKIOELEA 381
 Db 1196 KHRDLTFNHESLLRQSASYKEXKLSLASSENKDLNKKVSLTKQVNLSPKASKVPPEL 1255
 QY 382 R 382
 Db 1256 K 1256

RESULT 6
 TPR_HUMAN
 ID TPR_HUMAN STANDARD; PRT; 2349 AA.
 AC P12270;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NUCLEOPROTEIN TPR.
 GN TPR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93064711; PubMed=1437155;
 RA Mitchell P.J., Cooper C.S.;
 RT "The human tpr gene encodes a protein of 2094 amino acids that has
 extensive coiled-coil regions and an acidic C-terminal domain";
 RL Oncogene 7:2329-2333(1992).
 RN [2]
 RP REVISIONS, AND CHARACTERIZATION.
 RX MEDLINE=95096166; PubMed=7798308;
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
 RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
 RT "tpr, a large coiled coil protein whose amino terminus is involved in
 activation of oncogenic kinases, is localized to the cytoplasmic
 surface of the nuclear pore complex";
 RL J. Cell Biol. 127:1515-1526(1994).
 RN [3]
 RP SEQUENCE OF 1-142 FROM N.A.
 RX MEDLINE=88262257; PubMed=3387099;
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
 RT "tpr homologues activate met and raf";
 RL Oncogene 2:617-619(1988).
 CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
 COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
 IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
 COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
 TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
 COMPONENTS, INCLUDING P62.
 CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
 BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
 CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
 OR RAF GENES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X66397; CAA47021.1; -;
 DR EMBL; Y00672; CAA68681.1; -;
 DR PIR; S00928; S00928.
 DR MIM; 189940; -;
 KW Heptad repeat pattern; Coiled coil; Proto-oncogene;
 KW Chromosomal translocation; Nuclear protein; Transport.
 FT DOMAIN 78 360
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 422 571
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 575 628
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 758 805
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 834 869
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 934 979
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1004 1064
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1138 1166
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1196 1241
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1262 1304
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1354 1434
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1476 1595
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 527 530
 FT POLY-SER.
 FT DOMAIN 1833 1836
 FT POLY-GLU.
 FT DOMAIN 1957 1964
 FT POLY-ASP.
 FT DOMAIN 2295 2298
 FT POLY-SER.
 SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 10.6%; Score 206.5; DB 1; Length 2349;
 Best Local Similarity 21.4%; Pred. No. 0.013;
 Matches 102; Conservative 80; Mismatches 174; Indels 121; Gaps 17;
 QY 5 AFRSDTDTLEKVOER---ADKFEIENNTLKLNKSLDNFNKALKDHNDLTELSSNAKE 60
 Db 91 ALTEKNKELETAQDNIAIQSQFTFKEELEAKRDLRTNRELSSOELEYTEDVKRLNE 150
 QY 61 KLRGSVTRGTINDPQRAKEALDYELNHDKTKNEGLKTENEGKLTENEGKLTENEG 120
 Db 151 KLKESNTTKGEL-----QLKLDELQASDVSVKVRKLEQEKELHSHQNTWLATE---L 201
 QY 121 KTEVDVRVPRG-----TVENP---DKARELLNKYDVNSMLQANNDKLPWRV 164
 Db 202 KTKTDELLALGREGNEILELKNLENKKEEVSRLQEQMNGKLTSNELHOKHVEDLLTKL 261
 QY 165 RYTRH---TPEDKL-----KKTIIDDLDAKEHELOOQNEKLSLQNGDNPRE 207
 Db 262 KEAKEQQASMEKPHNELNAHLKSLNLYKSAADSEAKSNELTRAVEEL-----HK 312
 QY 208 VIEDLAANPAIONIRLR-HENKD-----LKALENAMVAGRDFKRAGTLLD 254
 Db 313 LLKEAGEANKAIQDHLEVEQSKDQMEKLEKIGRLEKELENANDLLSAT-KRKGAIS 371
 QY 255 Q-----VTOLYTKH-----NSNYQVYNAQAGRDL 279
 Db 372 EELAAAMSPATAAAVAKIVKPGMKITELYNVETQDQLLEKLENKINKIYLDIVK-EV 430
 QY 280 RQKAEYLGNDWAERLLQEL-NIDVATRSQDTLEKQVPRADKFEIENNTLKLNKSLDLS 338
 Db 431 EAKAPILKRRQREVEERAQKAVASLSVKLEQAMKEIQRLQEDTDKANKQSSVLERDNRME 490
 QY 339 FNNKALDHNDDELTEELSNK-----EKLKNDKSLSEKA-----SKIOELE 380
 Db 491 IOVKDLSQOIRVLLMELEFEARGNHVIRDEEVSSADISSSEVSIQHLVSYRNIELO 547

RESULT 7
 MYSACAEEL
 ID MYSACAEEL STANDARD; PRT; 1969 AA.
 AC P12844;

01-OCT-1989 (Rel. 12, Created)
01-OCT-1989 (Rel. 12, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN HEAVY CHAIN A (MHC A).
GN MYO-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=89178677; PubMed=2926820;
RA Dobb N.J., Maruyama I.N., Krause M., Karn J.:
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
heavy chain gene family."
RL J. Mol. Biol. 205:603-613(1989).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PPM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.
ELEGANS.
CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC -1- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X08067; CAA30856.1; -
PIR; S02771; S02771.
DR HSSP; P08799; 1MND.
DR InterPro; IPR000048; -
DR InterPro; IPR001609; -
DR InterPro; IPR002928; -
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
KW Multigene family.
FT DOMAIN 1 851 GLOBULAR HEAD (S1).
FT DOMAIN 852 1969 RODLIKE TAIL (S2 AND LMN DOMAINS).
FT DOMAIN 857 1969 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP.
FT DOMAIN 667 689 ACTIN-BINDING.
FT DOMAIN 770 784 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 717 717 ALKYLATION (SH-2) (POTENTIAL).
SEQUENCE 1969 AA; 225509 MW; 64577BBAF7EAD80A CRC64;

Query Match 10.4%; Score 203; DB 1; Length 1969;
Best Local Similarity 19.2%; Pred. No. 0.016;
Matches 101; Conservative 96; Mismatches 165; Indels 164; Gaps 18;
QY 1 ACWVATRSQDTDTLEKVRQADKFEIENNTLKLKNSDLSFNKALKDHNDLT-----E 53
DB 899 ALFSLSETEKANLADAEERNEKLNOLKATLESKLSLDITGOLEDQERNEDLARQKKTDQ 958
QY 54 ELSNAKEKLRSVTRGTINDPQRAKEALDKYLEHDLKTKNEGLTKNEGL----- 106
DB 959 ELSDTKKHVQDLEL---SLRKAQEQKSRD-----HNIRSLQDEMANQDAVAKLNKEK 1009
QY 107 KTEENGLKTENGLATEVDV-----PPRGTVENPDK--- 138
DB 1010 KHOESNRKLNDLOSEDKVNHLEKIRNKLEQQQWDELEENIDREKSRGDIETAKRKE 1069
QY 139 -----ARELLN---KYDVENSIMQANNDKLPWRVRYTRHTPD-----KLKKII 179
DB 1070 GDLKVAQENIDITKQKHQDVETTLARKEE-----LHHTNAKLAENNSIAKLQRLI 1121
QY 180 DDLDAK-----EHELOQNEKLSLQNG-----DGNP--- 205
DB 1122 KELTSARNAELEEAEARNRSDRSRSEARELEELTERLEQQGGATAAQAANKKRE 1181
QY 206 -----REVIEDLAANPAIONIRLHEN-----KOLKARLENAMVAGRDF 246
DB 1182 AETAKLRKEEDSLNHETATSSLRKRHGDVSVALETQLETQLKLUKAKSEAKSKLQKDL 1241
QY 247 -----KRAGTLLDQVTLQYTK---HNSNYQQYNAQAGRLDLQAQAEY 285
DB 1242 EESQATDSEVSRQDLEKALKATIEVOYSELQTKADEQSRQLODFAALKNRLN-NENSD 1300
QY 286 LKGLNDWAERLLOELNIDVATRSQDITLKVQERADKFEIENNTLKLKNSDLSFNKALK 345
DB 1301 NRSLEEMDNQLNSHLKSLTQSQ---LDETRNYDEESRERQALAAATAKLNHEHTILR 1357
QY 346 DINDELTELSNAKEKLREKDKSL-----SEKASKIOLEARK 383
DB 1358 EHLDEAESKADLTQISKLNAEQWKAQFDSGLNKLLEIEAAK 1403
RESULT 8
REST_HUMAN
ID REST_HUMAN STANDARD; PRT; 1427 AA.
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RESTIN (CYTOPLASMIC LINKER PROTEIN-170 ALPHA-2) (CLIP-170) (REED-
DE STERNBERG INTERMEDIATE FILAMENT ASSOCIATED PROTEIN).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood monocytes;
RX MEDLINE=92289675; PubMed=1600942;
RA Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M.,
RA Cerletti N., Sorg C., Odink K., Tarsay L., Wiesendanger W.,
RA de Wolf-Peters C., Shipman R.;
RT "Restin: a novel intermediate filament-associated protein highly
expressed in the Reed-Sternberg cells of Hodgkin's disease."
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92405160; PubMed=1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT "CLIP-170 links endocytic vesicles to microtubules."
RL Cell 70:887-900(1992).

CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64838; CAA46050.1; -
CC EMBL; M97501; AAA35693.1; -
CC PIR; S22695; S22695.
CC MIM; 179838; -
CC InterPro; IPR000938; -
CC Pfam; PF01302; CAP_GLY; 2.
CC PROSITE; PS00845; CAP_GLY_1; 2.
CC Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
KW DOMAIN 60 125 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT FT DOMAIN 214 279 CAP-GLY 2.
FT FT DOMAIN 304 331 SER-RICH.
FT FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
FT FT DOMAIN 1408 1421 CCHC-BOX.
FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
FT CONFLICT 1069 1069 D -> E (IN REF. 2).
SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match 10.3%; Score 202; DB 1; Length 1427;
Best Local Similarity 21.4%; Pred. No. 0.012;
Matches 95; Conservative 86; Mismatches 161; Indels 102; Gaps 18;

Qy 7 RSQDTLLEKQVRAKDFEINNTLK-----LNKSDLSFNKALKHNDLTELSENAKEK 61
Db 861 RSMQETVNLKQKEQFMNLSDDLEKLENLADMEAKFREK-----DEREQLIKAKEK 914
Qy 62 L-----RGSVTRGTINDPQRAKEALDKYLENHDLYKNEGLKTENGLKT 108
Db 915 LNDIAETIMKSGDSSQLTK--MNDELRLKER-----DVEELQLK-----LTK 956
Qy 109 ENEG---LKTENEGKTEVDVPRGTVNPDKARELLNKY-DVENSMLQANDKLPRV 164
Db 957 ANENASFLOKSIEDWTKAEQSQEAAKHEEKEKLERKLSLEKMKETSHNQOELKA 1016
Qy 165 RYRHTPEDKIL--KIIDDL-----AKE-----HELQOONKELSLQNG 201
Db 1017 RYERATSETKTHEILQNLQKTLTDTEKLGKAREENSGLLQLEELRQADKAAQAQT 1076
Qy 202 DGNPREVIEDLAANPAQNTLRHKNKOLKARLENAMVAGRDFKRACTLLDQVTLTYT 261
Db 1077 AEDAMQIMEQTKK--TETLASIEDTKQTNKQLQNELDTLKENNKNVBEELNKRSELLT 1134
Qy 262 KHNSNYQQYNAQAAGRLD--LRQAEYKLGNDWAERLLQELINIDVATRSQDTDLKQVQR 319
Db 1135 VENQKMEFREIEITLQAAQAKSQQLSALQENNVKLAELG---RSRDEVTSHQKLEE- 1190
Qy 320 ADKFEIENNTLKNLSDFNKKALKOHND-----LT-----EELSNK 359
Db 1191 -ERSVLNQLLEMKRE-----SKFIKDADEKASIQKSISITSALLTEKDAEKLKRENEV 1245
Qy 360 EKLKNDKSLSEKASKIQELEARK 383
Db 1246 TVLRGENASAKSLHVVQTLSEDK 1269

RESULT 9
MYSC_CHICK
ID MYSC_CHICK STANDARD; PRT; 1102 AA.
AC P29616;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ISOFORM (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 65-1102 FROM N.A.
RX STRAIN=BOILER BREEDERS; TISSUE=Heart;
RC MEDLINE=92130260; PubMed=1774788;
RA Stewart A.F.R., Camoretti-Mercado B., Perlman D., Gupta M.,
RT Jakovcic S., Zak R.;
RT "Structural and phylogenetic analysis of the chicken ventricular
RT myosin heavy chain rod";
RL J. Mol. Evol. 33:357-366(1991).
RN [2]
RP SEQUENCE OF 1-259.
RC TISSUE=Heart;
RX MEDLINE=93039740; PubMed=1418675;
RA Watanabe B.;
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT cardiac muscle myosin";
RL Biol. Chem. Hoppe-Seyler 373:1045-1054(1992).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS OF THE MYOFIBRILS.
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- TISSUE SPECIFICITY: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- DEVELOPMENT AND REGENERATION FOLLOWING FREEZE INJURY.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59552; CAA42130.1; -
CC HSP; P03069; IZIK.
CC InterPro; IPR002928; -
DR Pfam; PF01576; Myosin_tail; 1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Heptad repeat pattern; Multigene family.
FT NON_TER 1 1
FT DOMAIN <1 1102 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN <1 1102 COILED COIL (POTENTIAL).
FT CONFLICT 65 65 A -> T (IN REF. 1).
FT CONFLICT 92 92 V -> A (IN REF. 1).
FT CONFLICT 113 114 CS -> FAL (IN REF. 1).
FT CONFLICT 135 135 H -> Q (IN REF. 1).
FT CONFLICT 185 185 T -> R (IN REF. 1).
SQ SEQUENCE 1102 AA; 128008 MW; 2293668D049825DC CRC64;

SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;

 Query Match 10.2%; Score 200.; DB 1; Length 1208;
 Best Local Similarity 21.6%; Pred. No. 0.013;
 Matches 89; Conservative 77; Mismatches 160; Indels 86; Gaps

QY	3	MVATRSOTDTLEKVOERADKFIEINNTIKLNSDLSFNKALKDHNDDELTEELSNAK---	59
DB	427	MLQFSEEEEDVLSEKLQTLEDNNSLRLTWSLGNQIESLRITQNREIDEERNHLRLLA	486
QY	60	EKLGRSAVTGRTINDPQAKALDKYLEHNDLKTKNEGLKTENEGTKTENEGLTKEG	119
DB	487	SKNSDKAAETNIRLQEVTK-----ELET--LRMKNSDLNETHDURENEGILTKIDS	538
QY	120	LKTEVDVRFPRGTGVENPDKARELINLKDVENSMLQAANDKLPRWRYVTRHTPDCLKKII	179
DB	539	ITKEKDRLI-----NELEQRILKSVEVNSELNGTIDE----YRNKLKDKRETNEVM	586
QY	180	DLDLDAKHEYLQQOENEKLSLONGDNPREVIDEIAANPAIQNTLRHENKDLKARLENAM	239
DB	587	NAFOYKONDLRRFPESIN-----KLODREKELTSNLEKKN	621
QY	240	EVAGRDFKRAGTLDDQVTLQYTKHNSNYQYNQAQGRGLDLRQKAEVYLKGLNDWAERLQE	299
DB	622	LV-----ISSLRETYAMLEKERESIKKLYSGNARDLD---NTNMEILNDKISVLQRQ	671
QY	300	L-----NIDVATRSQTTLEKVQERADKFEI---ENNWLKLNKSDLS---FNNKALKDHN	348
DB	672	LTDVKDELDVSEEREBAIVAGOKLSAFELMSNEKQALEKYSSLKNELINAQNLDDR	731
QY	349	DELTEEI-----SNAKEKLKRN-----DKSLSEKASKIOELEARK	383
DB	732	EEELSLSKKLFBEERTKSRSGNDDIEKNVINVELADKLQAIRLES DK	783

```

RESULT 11
MYS_A_HUMAN
ID MYSA_HUMAN STANDARD; PRT; 1939 AA.
AC PI3533; Q13943; Q14906; Q14907;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DEF MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM.
GN MYH6 OR MYHCA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92133665; PubMed=1776652;
RA Matsuko R., Beisel K.W., Furutani M., Arai S., Takao A.;
RT "Complete sequence of human cardiac alpha-myosin heavy chain gene and
RT amino acid comparison to other myosins based on structural and
RT functional differences.";
RL Am. J. Med. Genet. 41:537-547(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94140346; PubMed=8307559;
RA Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;
RT "Structural organization of the human cardiac alpha-myosin heavy
RT chain gene (MYH6).";
RL Genomics 18:505-509(1993).
RN [3]
RP SEQUENCE OF 1-177 AND 1551-1939 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;

```



```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M76598; AAA37159.1; -
CC EMBL; M76599; AAA37160.1; -
CC EMBL; M76600; AAA37161.1; -
CC EMBL; M76601; AAA37162.1; -
CC EMBL; M62404; AAA37424.1; -
CC HSSP; P08799; 1MND.
CC SWISS-2DPAGE; Q02566; MOUSE.
CC MGD; MGI:97255; Myhca.
CC InterPro; IPR000048; -
CC InterPro; IPR001609; -
CC InterPro; IPR002928; -
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
CC Multigene family; Calmodulin-binding; Polymorphism.
CC DOMAIN 1 841
CC FT DOMAIN 842 1938
CC FT DOMAIN 842 1938
CC FT NP_BIND 178 185
CC FT DOMAIN 657 679
CC FT DOMAIN 759 773
CC FT DOMAIN 790 807
CC FT DOMAIN 816 833
CC FT MOD_RES 129 129
CC FT MOD_RES 697 697
CC FT MOD_RES 707 707
CC FT VARIANT 194 194
CC FT VARIANT 545 545
CC FT VARIANT 838 838
CC FT SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
KW Multigene family; Calmodulin-binding; Polymorphism.
FT DOMAIN 1 841
FT FT DOMAIN 842 1938
FT FT DOMAIN 842 1938
FT FT NP_BIND 178 185
FT FT DOMAIN 657 679
FT FT DOMAIN 759 773
FT FT DOMAIN 790 807
FT FT DOMAIN 816 833
FT FT MOD_RES 129 129
FT FT MOD_RES 697 697
FT FT MOD_RES 707 707
FT FT VARIANT 194 194
FT FT VARIANT 545 545
FT FT VARIANT 838 838
FT FT SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;
SQ
Query Match 10.0%; Score 195; DB 1; Length 1938;
Best Local Similarity 21.9%; Pred. No. 0.037;
Matches 97; Conservative 70; Mismatches 141; Indels 134; Gaps 18;
Qy 7 RSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELTELSNAKEKLRS- 65
Db 894 QAEQDNLDAAEECDQLIKKKIQLEAKVKEMT---EKLED-EEMNAELTAKARKLEDEC 949
Qy 66 AVTRGTINDPQRAKEALDYELNHDILKTNEGKLTENEGL-----KTENEGLK 114
Db 950 SELKKDIDD---LELTAKVKEKHATENKVNLTREMAGLDEIIAKLTKEKKALQEAHQ 1006
Qy 115 TENEGKTEVDVRVP-----RGIVENPKAKRELLNKYD-----VEN 150
Db 1007 QALDDLOAEEDKVNTLTAKSKVLEQQVDDLEGLSEQEKVKYRMDLERAKRKLGLDKLTQE 1066
Qy 151 SMLQANNDKLPWRYRTHPTEDLKKIIDDLDAKEHELOQQNEKL-----SL 198
Db 1067 SIMDLENDKLOL-----EELKK-----KEFDISQQNSKIEDEQALALQLOKKL 1110
Qy 199 QNGDGNPREVIEDLAANPA---TONTR--LRHENKDLKARLENAMVAGRDFKRACTLL 253
Db 1111 KENQARIEELEEELEAERTAKRVEKRLSRLSLELETSEERLEEA-----GGATS 1160
Qy 254 DQVTOLYTKHNSYQQYNAQAGRLDLKQAEYKGLNDWAERLLQELNIDVATR----- 307
Db 1161 VOI-----EMNKKRAEFQMKRRDLEATLQHEATAAALRRKKHADS 1201
```

```
Qy 308 -----SQDPTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELTELSNAKEKL 362
Db 1202 VAEIGEIQIDNLRVKQKLER---EKSEFKLELDDVTSN-----MEQIIKAKANL 1247
Qy 363 RKNDKSLSEKAS-----KIQELE 380
Db 1248 EKVSRTLEDOANEYRVKLEEAQ 1269
Search completed: May 10, 2001, 08:40:26
Job time: 154 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:38 ; Search time 27.19 Seconds
(without alignments)
1650.998 Million cell

Title: US-09-151-409-16
 Perfect score: 1954
 Sequence: 1 ACMVATRSOTDTLEKVOERA.....KNDKSLSKASKIOELEAR 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:	374700
--	--------

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```

1:  sp.archaea.*
2:  sp.bacteria.*
3:  sp.fungi.*
4:  sp.human.*
5:  sp.invertebrate.*
6:  sp.mammal.*
7:  sp.mbc.*
8:  sp.organelle.*
9:  sp.phage.*
10: sp.plant.*
11: sp.podent.*
12: sp.unclassified.*
13: sp.invertebrate.*
14: sp.virus.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	966	49.4	247	2	Q53344 streptococc
2	389	19.9	457	2	Q54510 streptococc
3	330	16.9	272	2	Q9KGR9 streptococc
4	330	16.9	539	2	Q54719 streptococc
5	321	16.4	355	2	Q10371 streptococc
6	320	16.4	581	2	Q54835 streptococc
7	317.5	16.2	231	2	Q9KGR8 streptococc
8	316	16.2	488	2	Q54830 streptococc
9	311	15.9	484	2	Q10372 streptococc
10	266.5	13.6	100	2	Q54632 streptococc
11	261.5	13.4	100	2	Q54635 streptococc
12	260.5	13.3	100	2	Q54638 streptococc
13	259.5	13.3	100	2	Q54637 streptococc
14	257	13.2	76	2	Q53536 streptococc
15	257	13.2	79	2	Q54543 streptococc
16	252	12.9	84	2	Q54598 streptococc
17	248	12.7	484	2	Q05464 streptococc
18	246	12.6	558	2	Q54718 streptococc
19	237	12.1	2167	10	Q9SSU1 Q9SSU1 chara coral

20	237	12.1	2182	10	Q9LW97
21	229	11.7	2138	5	Q9XZE3
22	228	11.7	83	2	Q9ZHE6
23	228	11.7	83	2	Q9ZHE6
24	226	11.6	2139	5	Q07569
25	227	11.6	1743	5	Q95063
26	220.5	11.3	327	2	Q9ZN08
27	214.5	11.0	150	2	Q54587
28	212.5	10.9	1945	6	Q37757
29	212.5	10.9	1526	3	Q9US16
30	212	10.8	1456	5	Q9V587
31	208.5	10.7	443	2	Q54703
32	207.5	10.6	709	4	Q14997
33	207.5	10.6	742	4	Q9UI01
34	207.5	10.6	980	5	Q96246
35	206.5	10.6	1950	5	Q26080
36	206.5	10.6	726	4	Q15624
37	206.5	10.6	1931	13	Q91973
38	205.5	10.5	2363	4	Q99968
39	204.5	10.5	1937	13	Q9IB04
40	204.5	10.5	911	3	Q06704
41	204	10.4	1927	5	Q25142
42	203	10.4	6677	5	Q9N435
43	203	10.4	1827	5	Q20042
44	203	10.4	1946	5	Q97291
45	202.5	10.4	1992	2	Q21440
46	202	10.4	1000	2	Q54633
47	202	10.4	1000	2	Q54633
48	202	10.4	1000	2	Q54633
49	202	10.4	1000	2	Q54633
50	202	10.4	1000	2	Q54633
51	202	10.4	1000	2	Q54633
52	202	10.4	1000	2	Q54633
53	202	10.4	1000	2	Q54633
54	202	10.4	1000	2	Q54633
55	202	10.4	1000	2	Q54633
56	202	10.4	1000	2	Q54633
57	202	10.4	1000	2	Q54633
58	202	10.4	1000	2	Q54633
59	202	10.4	1000	2	Q54633
60	202	10.4	1000	2	Q54633
61	202	10.4	1000	2	Q54633
62	202	10.4	1000	2	Q54633
63	202	10.4	1000	2	Q54633
64	202	10.4	1000	2	Q54633
65	202	10.4	1000	2	Q54633
66	202	10.4	1000	2	Q54633
67	202	10.4	1000	2	Q54633
68	202	10.4	1000	2	Q54633
69	202	10.4	1000	2	Q54633
70	202	10.4	1000	2	Q54633
71	202	10.4	1000	2	Q54633
72	202	10.4	1000	2	Q54633
73	202	10.4	1000	2	Q54633
74	202	10.4	1000	2	Q54633
75	202	10.4	1000	2	Q54633
76	202	10.4	1000	2	Q54633
77	202	10.4	1000	2	Q54633
78	202	10.4	1000	2	Q54633
79	202	10.4	1000	2	Q54633
80	202	10.4	1000	2	Q54633
81	202	10.4	1000	2	Q54633
82	202	10.4	1000	2	Q54633
83	202	10.4	1000	2	Q54633
84	202	10.4	1000	2	Q54633
85	202	10.4	1000	2	Q54633
86	202	10.4	1000	2	Q54633
87	202	10.4	1000	2	Q54633
88	202	10.4	1000	2	Q54633
89	202	10.4	1000	2	Q54633
90	202	10.4	1000	2	Q54633
91	202	10.4	1000	2	Q54633
92	202	10.4	1000	2	Q54633
93	202	10.4	1000	2	Q54633
94	202	10.4	1000	2	Q54633
95	202	10.4	10		

ALIGNMENTS

RESULT	1
Q53344	PRELIMINARY; PRT; 247 AA.
ID	Q53344
AC	Q53344;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE	TETRAVALENT M PROTEIN (FRAGMENT).
GN	EMM.
OS	Streptococcus sp.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1306;
[1]	SEQUENCE FROM N.A.
RN	MEDLINE=93346765; PubMed=8345202;
RX	Dale J.B., Chiang E.Y., Lederer J.W.;
RA	"Recombinant tetraivalent group A streptococcal M protein vaccine."
RT	J. Immunol. 151:2188-2194(1993).
RL	EMBL; S64396; AAB27745.1; -
DR	NCBI 247
FT	SEQUENCE 247 AA: 28252 MW: 7445827DE945529B CRC64:

Query Match	49.4%;	Score 966;	DB 2;	Length 247;
Best Local Similarity	78.9%;	Prod. No. 7.3e-39;		
Matches 195;	Conservative	0;	Mismatches 0;	Indels 52; Gaps 1;
3	Y	3	Y	3
1	B	1	B	1
63	Y	63	Y	63
61	B	61	B	61
71	Y	71	Y	71
121	B	121	B	121
131	Y	131	Y	131

Db 181 GTVNPDKARELLNKYDVNSMLQANDKLPWRVRYRHTPEDKLKKIIDDLDAKEHELQ 240

Qy 191 QQNEKLS 197
 |||||

Db 241 QQNEKLS 247

RESULT 2

Q54510 PRELIMINARY; PRT; 457 AA.

AC Q54510;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE M5.8193 PROTEIN (FRAGMENT).

GN EMM5.8193.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_taxID=1314;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC8193;

RC MEDLINE=94221454; PubMed=8170398;

RX Whatmore A.M., Kehoe M.A.;

RA "Horizontal gene transfer in the evolution of group A streptococcal

RT emm-like genes: gene mosaics and variation in vir regulons.";

RL MOL. Microbiol. 11:363-374(1994).

DR EMBL; U02480; AAA50854.1; -

DR INTERPRO; IPR001899; -

DR PFAM; PF00746; Gram_pos_anchor; 1.

DR PRINTS; PR00015; GPOSANCHOR.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

FT NON_TER 1

SQ SEQUENCE 457 AA; 51172 MW; DEBBAF550AEB6379 CRC64;

Query Match 19.9%; Score 389; DB 2; Length 457;

Best Local Similarity 34.4%; Pred. No. 1.7e-11;

Matches 123; Conservative 53; Mismatches 114; Indels 68; Gaps 14;

Qy 52 TEELSNAKEKLRGSAVIRGTINDPQRAKALDKYLENHLKTNNEGLKTENEGLKTENE 111
 ||||| : |||||

Db 13 TNEVS-----ATVTRGTINDPQRAKALDKYLENHLKTNNEGLKTENEGLKTENE 64
 ||||| : |||||

Qy 112 GLKTENEGLKTEVDPVPRGTVPDPKARELLNKYDVNSMLQANDKLPWRVRYRHTP 171
 ||||| ||||| : |||||

Db 65 GLKTENERLKTE-----KSNLER--KYAELTS-----EKKEHEAENDKLK--QQRDTLTSTQ 111
 ||||| : |||||

Qy 172 EDKLKKIIDDLDAKEHELQOONEKLSLQNGD-----GNPREVIDLAANPA 218
 ::||| : |||||

Db 112 KETLEREVQN-----TQYNNTLTKINGDLTKELNKTQELANKQOESKEKALNEL 164
 ||||| : |||||

Qy 219 IQ-----NIRLRHENKDLKARLENAMVAGRD-----FKRAGTLIDQVTLTYTHNS 265
 ||||| : |||||

Db 165 LKTVKDKIAKEQENKKTIGTLKKILDETVDKDLAKCQKSKQNTGALKQELAK---KDEA 221
 ||||| : |||||

Qy 266 NYOQYNAQAG-RIDLROKAEYLRKGLNDWAERLQLNELIDVATRS--OTDTIEKVQERADK 322
 ||||| : |||||

Db 222 NKIDSASRKGLRRDLDSAREAKKQLEAHQKLEBQNKISEASRRKGLRRD-LDASREAKQ 280
 ||||| : |||||

Qy 323 FETENNTLKLKNSDLSFNKALKDHNDDELTEELSNAKEKLRKNDKLSSEKASKIQELE 380
 ||||| : |||||

Db 281 LEAHQKLEBQNKISEASRRGLR-----LDASREAKKQVEKALEEANSKLALE 331
 ||||| : |||||

RESULT 3

Q9KGR9 PRELIMINARY; PRT; 272 AA.

AC Q9KGR9;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

Db 158 ATALEELEKKALELAIDQASODYNRANVLVELEETITREQENINLLGNAKLELDQ 217

QY 349 DELTEELSNAKEKLRK---NDKSLSEKASKTOELEAR 382

Db 218 SSEKQETIRLEKLNKELESKLTTERAK--AELQAK 252

RESULT 6

Q54835 ID Q54835 PRELIMINARY; PRT; 581 AA.

AC Q54835;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)

DE M3 PROTEIN PRECURSOR.

GN EMW3.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1314;

FP [1]

RN SEQUENCE FROM N.A.

RC STRAIN-TYPE M3 4 /55 (PRAGUE);

RA MEDLINE=96157064; PubMed=8586172;

RA Reichardt W., Gubbe K., Schmidt K.H.;

RT "M3-protein with close sequence homology to M12 protein binds

RT fibrinogen, albumin, fibronectin, but not to any subclass of IgG-

RT localization of binding regions";

RL Dev. Biol. Stand. 85:179-182 (1995).

DR EMBL; X80168; CAA56449.1; -

DR HSSP; P03069; IGCL.

DR INTERPRO; IPR001899; -

DR PFAM; PF00746; Gram_pos_anchor; 1.

DR PRINTS; PR00015; GPOSANCHOR.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

KW Signal.

FT SIGNAL 1 41 POTENTIAL.

FT CHAIN 42 581 M3 PROTEIN.

SQ SEQUENCE 581 AA; 64958 MW; 3689317C98359D75 CRC64;

Query Match 16.4%; Score 320; DB 2; Length 581;

Best Local Similarity 44.9%; Pred. No. 3.8e-08;

Matches 79; Conservative 21; Mismatches 36; Indels 40; Gaps

QY 234 RLENAMVAGRDFKRACTLLDQVLTQYTKNSNYQYNAQAGRLDRKAEYLKGLNDWA 293

Db 55 KLNKEIE-----NLLDQVTLKNSNYQYSAQAGRLDRKAEYLKGLNDWA 104

QY 294 ERLQELNIDVATR-----SQDTLLEK----VOERADKFEIENNTLKLNSDLSFNKA 343

Db 105 ERLQELNGEDVKKVLGKVAFKDDLEKEVKELKEIDKKEVEYQDLQ-KDFLAKQGYV 163

QY 344 LKD--HNDEL-----TELSNAKEKLKNDKSLSEKASKTOELE 360

Db 164 LSKDRHQOELEKEKKVTEATAKVGOISELETYVKOKVESTMODLTEKQNRYSQLE 219

RESULT 7

Q9KGR8 ID Q9KGR8 PRELIMINARY; PRT; 231 AA.

AC Q9KGR8;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE M PROTEIN (FRAGMENT).

GN EMW.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1314;

RN [1]

RESULT	7
O9KGR8	
ID	O9KGR8 PRELIMINARY; PRT; 231 AA.
AC	O9KGR8;
DT	01-OCT-2000 (TEMBLrel. 15, Created)
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT	01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE	M PROTEIN (FRAGMENT).
GN	EMM.
OS	Streptococcus pyogenes.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1314;
RN	[1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-7203-99;
RA Beall B.W.;
RT "5" emm sequence emm3.3.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF288753; AAF91497.1;
FT NON_TER 1
FT NON_TER 231
SQ SEQUENCE 231 AA; 26305 MW; F5002430DD229351 CRC64;

Query Match 16.2%; Score 317.5; DB 2; Length 231;
Best Local Similarity 44.8%; Pred. No. 1.9e-08;
Matches 81; Conservative 20; Mismatches 35; Indels 45; Gaps 7;

Qy 234 RLENAMEVAGDRFRAGTLLDQVTL-----YTKHNSYQYNAQAGRLDLRQKAEYLGK 288
   :||:|
Db 37 KLNKEIE-----NLLDQVTLTKHYTKHNSYQYNAQAGRLDLRQKAEYLGK 86

Qy 289 LNDWAERLQELNIDVATR-----SQTDTLEK----VOBRADKFEIENNTLKLKNSDLS 338
   |||||
Db 87 LNDWAERLQELNGEDVKVLGKVAPEKDDLEKEVKELKIDKKEKEYQDLD-KDFDLA 145

Qy 339 FNNKALKD--HNDEL-----TEESNAKEKLRKNDKSLSEKASKIQEEL 379
   |||
Db 146 KQGVLSDKRHOELEKEKKVTEATAKVGQISELETVKQKVESTMODLTEKQNRVSQL 205

Qy 380 E 380
Db 206 E 206

RESULT 8
Q54830 PRELIMINARY; PRT; 488 AA.
ID Q54830;
AC Q54830;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE M 3 PROTEIN PRECURSOR (FRAGMENT).
GN EM33.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_taxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1/29/58;
RX MEDLINE=93360826; PubMed=8355619;
RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
RT "Nucleotide substitutions and small-scale insertion produce size and
  antigenic variation in group A streptococcal M1 protein.";
RL Mol. Microbiol. 8:981-991(1993).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
  OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
  THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
  PHAGOCYTOSIS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
DR EMBL: X62131; CAA44062.1;
DR INTERPRO: IPR001899;
DR PFAM: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PRO0015; GPOSANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
KW Transmembrane; Coiled coil; Signal.
FT SIGNAL 1 41
FT CHAIN 42 484 M1.2 PROTEIN.
FT DOMAIN 42 458 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 134 244 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
FT DOMAIN 245 379 3 X APPROXIMATE TANDEM REPEATS, C-TYPE.
FT DOMAIN 413 449 GLY/PRO-RICH.
FT DOMAIN 450 455 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
  PROTEINS.
FT TRANSMEM 459 478 MEMBRANE ANCHOR.
FT DOMAIN 479 484 CYTOPLASMIC (POTENTIAL).
FT REPEAT 134 161 A-1.
FT REPEAT 162 189 A-2.
FT REPEAT 190 244 A-3.
FT REPEAT 245 286 C-1.
FT REPEAT 287 327 C-2.
FT REPEAT 328 379 C-3.
SQ SEQUENCE 484 AA; 54234 MW; 4179A7BBAC5FAED9 CRC64;

Query Match 15.9%; Score 311; DB 2; Length 484;
Best Local Similarity 44.2%; Pred. No. 8.3e-08;
Matches 96; Conservative 22; Mismatches 61; Indels 38; Gaps 9;

Qy 200 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGDRFKFA-----GTL 252
   |||||
Db 42 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGDRFKFAELEKAKAL 101

Qy 253 LDQVTLTKYKNSYQYNAQAGRLDLRQKAEYLGKGLNDWAERLQELNIDVATR--Q 309
   :||:|
Db 102 EDQKDLTKLQELQDYDLAKESTSWDRQRL--KELEKKEAL--ELAIQASRDYHR 157

```

FT	NON_TER	100	100
SQ	SEQUENCE	100 AA; 11248 MW; 76FE8A3C33C15CE2 CRC64;	
Query Match		13.4%; Score 261.5; DB 2; Length 100;	
Best Local Similarity		72.2%; Pred. No. 3.4e-06;	
Matches		57; Conservative	7; Gaps
OY	200	NGDGNPREVIEDLAANPAONTIRLRHENKDLKARLENAMEVAGRDFKRA-----GTL 252 	
Dd	16	NGDGSPREVIEDLAANPAONTIRLRHENKDLKARLENAMEVAGRDFKRAEELEKAQAL 75 	
OY	253	LDOVTQLTYTKHNSNYQQYN 271 	
Dd	76	EDQRKDLETKLKELQDDYD 94 	
RESULT 12			
O54638	ID	PRELIMINARY;	PRT; 100 AA.
AC	Q54638;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)		
DE	M1.6 PROTEIN (FRAGMENT).		
OS	Streptococcus pyogenes.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus		
OX	NCBI_TaxID=1314;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MGAS 2108;		
RX	MEDLINE=95172752; PubMed=7868273;		
RA	Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.:		
RT	"Genetic diversity and relationships among Streptococcus pyogenes		
RT	strains expressing serotype M1 protein: recent intercontinental spread		
RT	of a subclone causing episodes of invasive disease.";		
RL	Infect. Immun. 63:994-1003(1995).		
DR	EMBL; U20100; AAA85113.1; -;		
FT	NON_TER 1		
FT	NON_TER 100		
SQ	SEQUENCE 100 AA; 11301 MW; 63317108FC3A7DF2 CRC64;		
Query Match		13.3%; Score 260.5; DB 2; Length 100;	
Best Local Similarity		72.2%; Pred. No. 3.8e-06;	
Matches		57; Conservative	7; Gaps
OY	200	NGDGNPREVIEDLAANPAONTIRLRHENKDLKARLENAMEVAGRDFKRA-----GTL 255 	
Dd	16	NGDGSPREVIEDLAANPAONTIRLRHENKDLKARLENAMEVAGRDFKRAEELEKAQAL 75 	
OY	253	LDOVTQLTYTKHNSNYQQYN 271 	
Dd	76	EDQRKDLETKLKELQDDYD 94 	
RESULT 13			
O54637	ID	PRELIMINARY;	PRT; 100 AA.
AC	Q54637;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)		
DE	M1.5 PROTEIN (FRAGMENT).		
OS	Streptococcus pyogenes.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus		
OX	NCBI_TaxID=1314;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MGAS 1666;		
RX	MEDLINE=95172752; PubMed=7868273;		

RA Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.;
 RT "Genetic diversity and relationships among Streptococcus pyogenes
 strains expressing serotype M1 protein: recent intercontinental spread
 of a subclone causing episodes of invasive disease.";
 RL Infect. Immun. 63:994-1003(1995).
 DR EMBL; U20099; AAA85112.1; -;
 FT NON_TER 1
 FT NON_TER 100
 SQ SEQUENCE 100 AA; 11217 MW; 6332ACD6FFE7A3F2 CRC64;

Query Match 13.3%; Score 259.5; DB 2; Length 100;
 Best Local Similarity 72.2%; Pred. No. 4.2e-06;
 Matches 57; Conservative 1; Mismatches 14; Indels 7; Gaps 1;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMVAGRDFKRA-----GTL 252
 DB 16 NGDGNPREVIEGLAANNPAIONIRLRHENKDKLKALENAMVAGRDFKRAEEL 75
 QY 253 LDQVTQLYTKHNSNYQYN 271
 DB 76 EDQRKDLTKLKELOQDYD 94

RESULT 14

Q53536 PRELIMINARY; PRT; 76 AA.
 AC Q53536;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE M PROTEIN (FRAGMENT).
 GN EMM.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95377598; PubMed=7649435;
 RA Penney T.J., Martin D.R., Williams L.C., de Malmarche S.A.,
 RA Bergquist P.L.;
 RT "A single emm gene-specific oligonucleotide probe does not recognise
 all members of the Streptococcus pyogenes M type 1.";
 RL FEMS Microbiol. Lett. 130:145-149(1995).
 DR EMBL; S79254; AAB35162.1; -;
 FT NON_TER 1
 FT NON_TER 76
 SQ SEQUENCE 76 AA; 8383 MW; EA46B64E5657A220 CRC64;

Query Match 13.2%; Score 257; DB 2; Length 76;
 Best Local Similarity 96.2%; Pred. No. 4.1e-06;
 Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMVAGRDFKRACTL 252
 DB 22 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMVAGRDFKRAEEL 74

RESULT 15

Q54543 PRELIMINARY; PRT; 79 AA.
 AC Q54543;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE M TYPE 1 (EMML) (FRAGMENT).
 GN EMM.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=M TYPE 1;
 RA Whatmore A.M.;
 RL Thesis (1993), University of Newcastle Upon Tyne, UK.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M TYPE 1;
 RX MEDLINE=95198537; PubMed=7891551;
 RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
 RT "Non-congruent relationships between variation in emm gene sequences
 and the population genetic structure of group A streptococci.";
 RL Mol. Microbiol. 14:619-631(1994).
 DR EMBL; U11940; AAA95556.1; -;
 DR INTERPRO; IPR000886; -;
 DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 79
 SQ SEQUENCE 79 AA; 8499 MW; AE8A1FF8A4E1B711 CRC64;

Query Match 13.2%; Score 257; DB 2; Length 79;
 Best Local Similarity 96.2%; Pred. No. 4.3e-06;
 Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMVAGRDFKRACTL 252
 DB 27 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMVAGRDFKRAEEL 79

Search completed: May 10, 2001, 08:40:08
 Job time: 150 sec

us-09-151-409-16.rspt

Tue May 15 07:29:01 2001

